

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: July 18, 2004, 09:44:14 ; Search time 52 Seconds  
(without alignments)  
4465.796 Million cell updates/sec

Title: US-10-045-815-4  
Perfect score: 736  
Sequence: 1 MAVRALKLLTLLAVVAAS.....PPEPAQSLSGSEKPKDEL 736

Scoring table: OLIGO  
Gapop 60.0 , Gapext 60.0

Searched: 1017041 seqs, 315518202 residues

Word size : 0

Total number of hits satisfying chosen parameters: 1017041

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database : SPTREMBL 25: \*  
1: sp\_archaea: \*  
2: sp\_bacteria: \*  
3: sp\_fungi: \*  
4: sp\_human: \*  
5: sp\_invertebrate: \*  
6: sp\_mammal: \*  
7: sp\_mhc: \*  
8: sp\_organelle: \*  
9: sp\_phase: \*  
10: sp\_plant: \*  
11: sp\_rodent: \*  
12: sp\_virus: \*  
13: sp\_vertebrate: \*  
14: sp\_unclassified: \*  
15: sp\_rvirus: \*  
16: sp\_bacteriaph: \*  
17: sp\_archaeap: \*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES			
Result No.	Score	Query Match	Description
1	736	100.0	Q9HC86 homo sapien
2	433	58.8	Q96SK8
3	407	55.3	Q96SL5
4	360	48.9	Q9HC87
5	289	39.2	Q96BR8
6	281	38.2	Q96SN3
7	188	25.5	Q96HJ3
8	101	13.7	Q9RLJ8
9	101	13.7	Q9CFL3
10	101	13.7	Q9CWK5
11	91	12.4	Q9QZT7
12	82	11.1	Q9QZT6
13	80	10.9	Q9BSV2
14	16	2.2	Q8C673
15	16	2.2	Q8CG71
16	13	1.8	Q9NV12

17	13	1.8	708	4	Q8IVL5	Q8IVL5 homo sapien
18	10	1.4	412	5	Q9NDP6	Q9NDP6 ciona intes
19	8	1.1	133	8	Q36544	Q36544 scarabaeus
20	8	1.1	187	16	Q7V309	Q7V309 prochlorococ
21	8	1.1	196	16	Q8A8T9	Q8A8T9 bacteroides
22	8	1.1	222	16	Q9PP39	Q9PP39 campylobact
23	8	1.1	227	4	Q33513	Q33513 homo sapien
24	8	1.1	241	8	Q8WAQ0	Q8WAQ0 agabus bino
25	8	1.1	241	8	Q8WAQ2	Q8WAQ2 agabus bigu
26	8	1.1	241	8	Q8WSH1	Q8WSH1 agabus heyd
27	8	1.1	241	8	Q8WAP9	Q8WAP9 agabus dila
28	8	1.1	241	8	Q8WSH0	Q8WSH0 agabus bigu
29	8	1.1	241	8	Q958P2	Q958P2 agabus bipu
30	8	1.1	241	8	Q8WAQ1	Q8WAQ1 agabus alex
31	8	1.1	241	8	Q8WAQ3	Q8WAQ3 agabus gutt
32	8	1.1	241	8	Q8WAP8	Q8WAP8 agabus alex
33	8	1.1	241	8	Q8WAP7	Q8WAP7 agabus cf.
34	8	1.1	252	8	Q8HEV5	Q8HEV5 agabus disi
35	8	1.1	257	5	O16619	O16619 caenorhabdi
36	8	1.1	274	8	Q8SA66	Q8SA66 drosophila
37	8	1.1	288	10	Q84RS4	Q84RS4 oryza sativ
38	8	1.1	309	16	Q8PM96	Q8PM96 xanthomonas
39	8	1.1	317	16	Q9XAN8	Q9XAN8 streptomyce
40	8	1.1	388	4	Q8NB01	Q8NB01 homo sapien
41	8	1.1	443	11	Q8K2B0	Q8K2B0 mus musculu
42	8	1.1	452	16	Q8A4S2	Q8A4S2 bacteroides
43	8	1.1	457	11	Q8C1I2	Q8C1I2 mus musculu
44	8	1.1	466	8	Q9MT06	Q9MT06 ismene narc
45	8	1.1	493	12	Q993S9	Q993S9 anticarsia

ALIGNMENTS

RESULT 1

Q9HC86 PRELIMINARY; PRT; 736 AA.

ID Q9HC86

AC Q9HC86; (TREMREL. 16, Created)

DT 01-VAR-2001 (TREMREL. 16, Last sequence update)

DT 01-VAR-2001 (TREMREL. 16, Last sequence update)

DT 01-OCT-2003 (TREMREL. 25, Last annotation update)

DE GROS1-L protein.

OS Homo sapiens (Human).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

OX NCBI\_TaxID=9606;

RN [1]\_TaxID=9606;

RP SEQUENCE FROM N.A.

RC TISSUE=Testis;

RA MEDLINE=20406537; PubMed=10951563;

RA Kaul S.C., Sugihara T., Yoshida A., Nomura H., Wachwa R.;

RT "Gros1, a potential growth suppressor on chromosome 1: its identity to

RT basement membrane-associated proteoglycan, leprecan.";

RL Oncogene 19:3576-3583(2000).

DR EMBL; AF097432; AAC31019.1; --

DR GO; GO:0016706; F:oxidoreductase activity, acting on paired d. . .; IEA.

DR GO; GO:0019538; P:protein metabolism; IEA.

DR InterPro; IPR005123; 2OG-Fell Oxy.

DR InterPro; IPR000866; ER target S.

DR InterPro; IPR006620; Fco-4\_hyd\_alph.

DR InterPro; IPR008941; IPR-like.

DR Pfam; PF03171; 2OG-Fell Oxy; 1.

DR SMART; SM00702; P4HG; 1.

DR PROSITE; PS00014; ER TARGET; 1.

SQ SEQUENCE 736 AA; 83413 MW; 59D1205DB76ADB50 CRC64;

Query Match 100.0%; Score 736; DB 4; Length 736;

Best Local Similarity 100.0%; Pred. No. 0;

Matches 736; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MAVRALKLLTLLAVVAASQAQVESEAGWGMVTPDLLFAEGTAAYARGDWPGVLSMER 60

DB 1 MAVRALKLLTLLAVVAASQAQVESEAGWGMVTPDLLFAEGTAAYARGDWPGVLSMER 60

QY 61 ALRSRAALRALRLRCRTQCAADFPWELDPDWSPSPAQASGAGALRDLSPFGGLRRAACL 120  
DB 61 ALRSRAALRALRLRCRTQCAADFPWELDPDWSPSPAQASGAGALRDLSPFGGLRRAACL 120  
QY 121 RRCGLGPPAAHSLSEMELEFRKSPYNYLQVAFKINKLEKAVAAAHFTFFVGNPEHEMQ 180  
DB 121 RRCGLGPPAAHSLSEMELEFRKSPYNYLQVAFKINKLEKAVAAAHFTFFVGNPEHEMQ 180  
QY 181 QNLDDYQTMGSVKVADPKDLETPHMQEPRGLVRLYSEEQPQAVPHLEALQYFVAYE 240  
DB 181 QNLDDYQTMGSVKVADPKDLETPHMQEPRGLVRLYSEEQPQAVPHLEALQYFVAYE 240  
QY 241 ECRALCEGPDYDGYNYLEYNADLFOAITDHYIQVNLCKQNCVTTELASHPSREKPFDEL 300  
DB 241 ECRALCEGPDYDGYNYLEYNADLFOAITDHYIQVNLCKQNCVTTELASHPSREKPFDEL 300  
QY 301 PSHNYLQFAYNYIGNYTOAGECAKTYLLFPFNDENVNQNLAAYAAMLGEEHTRSIGPRE 360  
DB 301 PSHNYLQFAYNYIGNYTOAGECAKTYLLFPFNDENVNQNLAAYAAMLGEEHTRSIGPRE 360  
QY 361 SAKYRORSLEKELLFPAYDVFGIPFVDPDSWTPPEEVI PKRLOEKQKSERETAVRISQE 420  
DB 361 SAKYRORSLEKELLFPAYDVFGIPFVDPDSWTPPEEVI PKRLOEKQKSERETAVRISQE 420  
QY 421 IGMLMKEIETLVEEKTESLDVSRLTREGGPLYEGISLTWNSKLLNGYQVVMGVID 480  
DB 421 IGMLMKEIETLVEEKTESLDVSRLTREGGPLYEGISLTWNSKLLNGYQVVMGVID 480  
QY 481 HECQELQRLTNVAATSGDGRGQTSPTPNEKFGVTVFKALKLGQEGKVPLOSALHLYN 540  
DB 481 HECQELQRLTNVAATSGDGRGQTSPTPNEKFGVTVFKALKLGQEGKVPLOSALHLYN 540  
QY 541 VTEKVRIMESYFRDLTPPLFSYSHLVCRTAIEVQAEKDDSHPHVNDNCILNAETLVC 600  
DB 541 VTEKVRIMESYFRDLTPPLFSYSHLVCRTAIEVQAEKDDSHPHVNDNCILNAETLVC 600  
QY 601 VKEPPAYTFRDYSAILYNGDFDGNFYFTELDAKTVTAEVQCCGRAVGFSSGTENPHG 660  
DB 601 VKEPPAYTFRDYSAILYNGDFDGNFYFTELDAKTVTAEVQCCGRAVGFSSGTENPHG 660  
QY 661 VKAVTRGORCAIALWFTLDPHRSERDQVADDLVKMLFSPPEMDLSQEQPLDQAQGPPEP 720  
DB 661 VKAVTRGORCAIALWFTLDPHRSERDQVADDLVKMLFSPPEMDLSQEQPLDQAQGPPEP 720  
QY 721 AQESLSGSESKEPKDEL 736  
DB 721 AQESLSGSESKEPKDEL 736

## RESULT 2

Q96SK8 PRELIMINARY; PRT; 736 AA.  
AC Q96SK8  
DT 01-DEC-2001 (TrEMBLrel. 19, Created)  
DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)  
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)  
DE Hypothetical protein FLJ14791 (Hypothetical protein  
NT2P2005671).  
OS Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
OX NCBI\_taxID=9606;  
RN [1]  
RP SEQUENCE FROM N.A.  
RA Isogai T., Ota T., Hayashi K., Sugiyama T., Otsuki T., Suzuki Y.,  
RA Nishikawa T., Nagai K., Sugano S., Aotsuka S., Yoshikawa Y.,  
RA Matsumura H., Ishii S., Kawai Y., Saito K., Yamamoto J., Wakamatsu A.,  
RA Nakamura Y., Nagahari K., Masuho Y., Sasaki N.;  
RT "NSDO human cDNA sequencing project";  
RL Submitted (MAY-2001) to the ENBL/GenBank/DBJ databases.  
[2]  
RP SEQUENCE FROM N.A.

RA Ota T., Nishikawa T., Suzuki Y., Kawai-Hio Y., Hayashi K., Ishii S.,  
RA Saito K., Yamamoto J., Wakamatsu A., Nagai T., Nakamura Y.,  
RA Nagahari K., Sugano S., Isogai T.,  
RT "HRI human cDNA sequencing project";  
RL Submitted (MAR-2002) to the ENBL/GenBank/DBJ databases.  
DR EMBL; AK027697; BAB55305.1; -;  
DR EMBL; AK075418; BAC11608.1; -;  
DR InterPro; IPR005123; 2OG-Fell\_Oxy.  
DR InterPro; IPR000886; ER-target\_S.  
DR InterPro; IPR008941; TRP-like\_S.  
DR Pfam; PF03171; 2OG-Fell\_Oxy; 1.  
DR PROSITE; PS00014; ER\_TARGET; 1.  
KW Hypothetical protein.  
SQ SEQUENCE 736 AA; 83393 MW; EA1909828FAB685E CRC64;  
Query Match 58.8%; Score 433; DB 4; Length 736;  
Best Local Similarity 99.6%; Pred. No. 0;  
Matches 733; Conservative 0; Mismatches 3; Indels 0; Gaps 0;  
QY 1 MAVRALKLLTLLAVVAAASQAEESEAGWGKVTPLLFAEGTAAAYARGDPGVLSMER 60  
DB 1 MAVRALKLLTLLAVVAAASQAEESEAGWGKVTPLLFAEGTAAAYARGDPGVLSMER 60  
QY 61 ALRSRAALRALRLRCRTQCAADFPWELDPDWSPSPAQASGAGALRDLSPFGGLRRAACL 120  
DB 61 ALRSRAALRALRLRCRTQCAADFPWELDPDWSPSPAQASGAGALRDLSPFGGLRRAACL 120  
QY 121 RRCGLGPPAAHSLSEMELEFRKSPYNYLQVAFKINKLEKAVAAAHFTFFVGNPEHEMQ 180  
DB 121 RRCGLGPPAAHSLSEMELEFRKSPYNYLQVAFKINKLEKAVAAAHFTFFVGNPEHEMQ 180  
QY 181 QNLDDYQTMGSVKVADPKDLETPHMQEPRGLVRLYSEEQPQAVPHLEALQYFVAYE 240  
DB 181 QNLDDYQTMGSVKVADPKDLETPHMQEPRGLVRLYSEEQPQAVPHLEALQYFVAYE 240  
QY 241 ECRALCEGPDYDGYNYLEYNADLFOAITDHYIQVNLCKQNCVTTELASHPSREKPFDEL 300  
DB 241 ECRALCEGPDYDGYNYLEYNADLFOAITDHYIQVNLCKQNCVTTELASHPSREKPFDEL 300  
QY 301 PSHNYLQFAYNYIGNYTOAGECAKTYLLFPFNDENVNQNLAAYAAMLGEEHTRSIGPRE 360  
DB 301 PSHNYLQFAYNYIGNYTOAGECAKTYLLFPFNDENVNQNLAAYAAMLGEEHTRSIGPRE 360  
QY 361 SAKYRORSLEKELLFPAYDVFGIPFVDPDSWTPPEEVI PKRLOEKQKSERETAVRISQE 420  
DB 361 SAKYRORSLEKELLFPAYDVFGIPFVDPDSWTPPEEVI PKRLOEKQKSERETAVRISQE 420  
QY 421 IGMLMKEIETLVEEKTESLDVSRLTREGGPLYEGISLTWNSKLLNGYQVVMGVID 480  
DB 421 IGMLMKEIETLVEEKTESLDVSRLTREGGPLYEGISLTWNSKLLNGYQVVMGVID 480  
QY 481 HECQELQRLTNVAATSGDGRGQTSPTPNEKFGVTVFKALKLGQEGKVPLOSALHLYN 540  
DB 481 HECQELQRLTNVAATSGDGRGQTSPTPNEKFGVTVFKALKLGQEGKVPLOSALHLYN 540  
QY 541 VTEKVRIMESYFRDLTPPLFSYSHLVCRTAIEVQAEKDDSHPHVNDNCILNAETLVC 600  
DB 541 VTEKVRIMESYFRDLTPPLFSYSHLVCRTAIEVQAEKDDSHPHVNDNCILNAETLVC 600  
QY 601 VKEPPAYTFRDYSAILYNGDFDGNFYFTELDAKTVTAEVQCCGRAVGFSSGTENPHG 660  
DB 601 VKEPPAYTFRDYSAILYNGDFDGNFYFTELDAKTVTAEVQCCGRAVGFSSGTENPHG 660  
QY 661 VKAVTRGORCAIALWFTLDPHRSERDQVADDLVKMLFSPPEMDLSQEQPLDQAQGPPEP 720  
DB 661 VKAVTRGORCAIALWFTLDPHRSERDQVADDLVKMLFSPPEMDLSQEQPLDQAQGPPEP 720  
QY 721 AQESLSGSESKEPKDEL 736  
DB 721 AQESLSGSESKEPKDEL 736



```
Q96SL5
ID Q96SL5 PRELIMINARY; PRT; 736 AA.
AC Q96SL5;
DT 01-DEC-2001 (TrEMBLrel. 19, Created)
DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Hypothetical protein FLJ14774.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
CX NCBI_TaxID=9606;
RN [1]
SEQUENCE FROM N.A.
RA Isogai T., Ota T., Hayashi K., Sugiyama T., Otsuki T., Suzuki Y.,
RA Nishikawa T., Nagai K., Sugano S., Aotsuka S., Yoshikawa Y.,
RA Matsunawa H., Ishii S., Kawai Y., Saito K., Yamamoto J., Wakamatsu A.,
RA Nakamura Y., Nagahari K., Masuho Y., Sasaki N.;
RT "NEDO human cDNA sequencing project.";
RL Submitted (MAY-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; AK027680; BAB55291.1; -.
DR GO; GO:0016706; F:oxidoreductase activity, acting on paired d. ...; IEA.
DR GO; GO:0019538; P:protein metabolism; IEA.
DR InterPro; IPR005123; 2OG-Fell_Oxy.
DR InterPro; IPR000886; ER_Farset_Oxy.
DR InterPro; IPR006620; Pro_4_hyd_alph.
DR InterPro; IPR008941; TPR-like.
DR Pfam; PF03171; 2OG-Fell_Oxy; 1.
DR SMART; SM00702; P4HC; 1.
DR PROSITE; PS00014; ER_TARGET; 1.
KW Hypothetical protein.
SQ
SEQUENCE 736 AA; 83411 MW; EA1819938PAE685E CRC64;

Query Match 55.3%; Score 407; DB 4; Length 736;
Best Local Similarity 99.8%; Pred. No. 0;
Matches 707; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 MAVRALKLLTLLAVAAASQAQVESEAGWGVTDPDLLFAEGTAAYARGDWPVLSMER 60
DB 1 MAVRALKLLTLLAVAAASQAQVESEAGWGVTDPDLLFAEGTAAYARGDWPVLSMER 60
QY 61 ALRSRAALRALRLRCRTQCAADFPWELDPDWSPPAQASGAGALRDLSPFGGLLRAACL 120
DB 61 ALRSRAALRALRLRCRTQCAADFPWELDPDWSPPAQASGAGALRDLSPFGGLLRAACL 120
QY 121 RRCGLGPPAAHSLSEEMLEFRKSPNYLQVAFKINKLEKAVAAAHFFVGNPEHMEQ 180
DB 121 RRCGLGPPAAHSLSEEMLEFRKSPNYLQVAFKINKLEKAVAAAHFFVGNPEHMEQ 180
QY 181 QNLDDYQTMGSKVEADFKDLETQPHMQEFLGLVRLYSEEQPQEAAPHLEAALQEFVAYE 240
DB 181 QNLDDYQTMGSKVEADFKDLETQPHMQEFLGLVRLYSEEQPQEAAPHLEAALQEFVAYE 240
QY 241 ECRALCEGPDYDGYNYLQVAFKINKLEKAVAAAHFFVGNPEHMEQ 300
DB 241 ECRALCEGPDYDGYNYLQVAFKINKLEKAVAAAHFFVGNPEHMEQ 300
QY 361 SAKERYQRSLLEKELLFFAYDVGIPFVDPDPSWTPEEVIKPKLQEKQSERETAVRISQE 420
DB 361 SAKERYQRSLLEKELLFFAYDVGIPFVDPDPSWTPEEVIKPKLQEKQSERETAVRISQE 420
QY 421 IGNLMKEITLVBEKTKESLDYSRLTREGGPLLLEGISLTWNSKLLNGVQVMDGVISD 480
DB 421 IGNLMKEITLVBEKTKESLDYSRLTREGGPLLLEGISLTWNSKLLNGVQVMDGVISD 480
QY 481 HFCQBLQRLTNVAATSGDGYRGQTSPHTPNEKFGYGVTVFKALKLQEGKVPQLSAHLYYN 540
DB 481 HFCQBLQRLTNVAATSGDGYRGQTSPHTPNEKFGYGVTVFKALKLQEGKVPQLSAHLYYN 540
QY 541 VTEKVRIMESYFRLDTPLYFSYSHLVCTAEEVQAEKDDSHPHVYDNCILNAETLVC 600
```

```
DB 541 VTEKVRIMESYFRLDTPLYFSYSHLVCTAEEVQAEKDDSHPHVYDNCILNAETLVC 600
QY 601 VKEPPAYTFRDYSAILYNGDPDGNFYFTELDATKVTAEVQPCQGRAVGFSSGTENPHG 660
DB 601 VKEPPAYTFRDYSAILYNGDPDGNFYFTELDATKVTAEVQPCQGRAVGFSSGTENPHG 660
QY 661 VKAVTRGRCALALWFTLDPHRSERDRVQADDLVKMLFSPREMDSLQSQP 710
DB 661 VKAVTRGRCALALWFTLDPHRSERDRVQADDLVKMLFSPREMDSLQSQP 710

RESULT 4
Q9HC87
ID Q9HC87 PRELIMINARY; PRT; 363 AA.
AC Q9HC87;
DT 01-WAR-2001 (TrEMBLrel. 16, Created)
DT 01-WAR-2001 (TrEMBLrel. 16, Last sequence update)
DE GROSI-S protein.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
CX NCBI_TaxID=9606;
RN [1]
SEQUENCE FROM N.A.
RC TISSUE=Testis;
RX MEDLINE=20406537; PubMed=10951563;
RA Kaul S.C., Sugihara T., Yoshida A., Nomura H., Wadhwa R.;
RT "Grosi, a potential growth suppressor on chromosome 1: its identity to
RT basement membrane-associated proteoglycan, leprecan.";
RL OncoGene 19:3576-3583(2000).
DR EMBL; AF097431; AAG31018.1; -.
DR GO; GO:0008285; P:negative regulation of cell proliferation; NAS.
DR InterPro; IPR008941; TPR-like.
SQ
SEQUENCE 363 AA; 41158 MW; 5B1B66B38679E76A CRC64;

Query Match 48.9%; Score 360; DB 4; Length 363;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 360; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MAVRALKLLTLLAVAAASQAQVESEAGWGVTDPDLLFAEGTAAYARGDWPVLSMER 60
DB 1 MAVRALKLLTLLAVAAASQAQVESEAGWGVTDPDLLFAEGTAAYARGDWPVLSMER 60
QY 61 ALRSRAALRALRLRCRTQCAADFPWELDPDWSPPAQASGAGALRDLSPFGGLLRAACL 120
DB 61 ALRSRAALRALRLRCRTQCAADFPWELDPDWSPPAQASGAGALRDLSPFGGLLRAACL 120
QY 121 RRCGLGPPAAHSLSEEMLEFRKSPNYLQVAFKINKLEKAVAAAHFFVGNPEHMEQ 180
DB 121 RRCGLGPPAAHSLSEEMLEFRKSPNYLQVAFKINKLEKAVAAAHFFVGNPEHMEQ 180
QY 181 QNLDDYQTMGSKVEADFKDLETQPHMQEFLGLVRLYSEEQPQEAAPHLEAALQEFVAYE 240
DB 181 QNLDDYQTMGSKVEADFKDLETQPHMQEFLGLVRLYSEEQPQEAAPHLEAALQEFVAYE 240
QY 241 ECRALCEGPDYDGYNYLQVAFKINKLEKAVAAAHFFVGNPEHMEQ 300
DB 241 ECRALCEGPDYDGYNYLQVAFKINKLEKAVAAAHFFVGNPEHMEQ 300
QY 301 PSHNYLQFAYNIGNYTOAGECAKTYLLFFPNDEVNQNLAAYAAMLGEEHTRSIGPRE 360
DB 301 PSHNYLQFAYNIGNYTOAGECAKTYLLFFPNDEVNQNLAAYAAMLGEEHTRSIGPRE 360

RESULT 5
Q96BR8
ID Q96BR8 PRELIMINARY; PRT; 390 AA.
AC Q96BR8;
DT 01-DEC-2001 (TrEMBLrel. 19, Created)
DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
```

DE Growth suppressor 1 (Leucine proline-enriched proteoglycan (Leprecan)  
DE 1).  
DE Homo sapiens (Human).  
OS Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
OC NCBI\_TaxID=9606;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC TISSUE=Ductenium;  
RA Strausberg R.;  
RL Submitted (OCT-2001) to the EMBL/GenBank/DBJ databases.  
RN [2]  
RN SEQUENCE FROM N.A.  
RP Kaline N., Chen X., Rolfs A., Halleck A., Hines L., Eisenstein S.,  
RA Koundinya M., Raphael J., Moreira D., Kelley T., LaBaer J., Lin Y.,  
RA Phelan M., Farmer A.;  
RT "Cloning of human full-length cDNAs in BD Creator(TM) System Donor  
vector.";  
RL Submitted (MAY-2003) to the EMBL/GenBank/DBJ databases.  
DR EMBL; BC015309; AAF15309.1; -.  
DR EMBL; BT007039; AAP35688.1; -.  
DR GO; GO:0016706; F:oxidoreductase activity, acting on paired d. . . ; IEA.  
DR GO; GO:0019538; P:protein metabolism; IEA.  
DR InterPro; IPR005123; 2OG-Fell\_Oxy.  
DR InterPro; IPR008886; ER\_target\_S.  
DR Pfam; PF03171; 2OG-Fell\_Oxy; 1.  
DR SMART; SMO0702; P4HC; 1.  
DR PROSITE; PS00014; ER\_TARGET; 1.  
SQ SEQUENCE 390 AA; 44077 MW; 61C9F5983CF637FB CRC64;  
Query Match 39.3%; Score 289; DB 4; Length 390;  
Best Local Similarity 99.7%; Pred. No. 8.4e-286;  
Matches 389; Conservative 0; Mismatches 1; Indels 0; Gaps 0;  
QY 347 MLGSEHRSIGPRESAKYRQSLLEKELFFAYDFVGFIPFVDPDSWTPPEVVKRLQEK 406  
DB 1 MLGSEHRSIGPRESAKYRQSLLEKELFFAYDFVGFIPFVDPDSWTPPEVVKRLQEK 60  
QY 407 QKSRETAIVSISQIGNLMKEITLVBEKTESLDVSRITREGGPIIYEGISLTWNSKLL 466  
DB 61 QKSRETAIVSISQIGNLMKEITLVBEKTESLDVSRITREGGPIIYEGISLTWNSKLL 120  
QY 467 NGYQRVMDGVI SDHCEQELRLTNVAATSGDGYRGOTSPHTNEKPYGVTVFKALKLGG 526  
DB 121 NGYQRVMDGVI SDHCEQELRLTNVAATSGDGYRGOTSPHTNEKPYGVTVFKALKLGG 180  
QY 527 EGKVPQASAHLYNNVTEKVRIMESYFRLDTPLYFSYSHLVCRTAIEVQAEKDDSHPV 586  
DB 181 EGKVPQASAHLYNNVTEKVRIMESYFRLDTPLYFSYSHLVCRTAIEVQAEKDDSHPV 240  
QY 587 HVDNCILNAETLVCKEPPATFRDYSAIYLNGDFDGGNFYFTELDKTVTAEVQPCG 646  
DB 241 HVDNCILNAETLVCKEPPATFRDYSAIYLNGDFDGGNFYFTELDKTVTAEVQPCG 300  
QY 647 RAVGFSSGTENPHGVKAVTRQRCRCAIALWFTLDPHRSERDRVQADDLVKMLFSPERM 706  
DB 301 RAVGFSSGTENPHGVKAVTRQRCRCAIALWFTLDPHRSERDRVQADDLVKMLFSPERM 360  
QY 707 QEQLDQAQGGPEPAQESLSGSESKPKDEL 736  
DB 361 QEQLDQAQGGPEPAQESLSGSESKPKDEL 390  
RESULT 6  
Q96SN3 PRELIMINARY; PRT; 804 AA.  
AC Q96SN3  
DT 01-DEC-2001 (TrEMBLrel. 19, Created)  
DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)  
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)  
DE Hypothetical protein FLJ14742.  
OS Homo sapiens (Human).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
OC NCBI\_TaxID=9606;  
RN [1]  
RP SEQUENCE FROM N.A.  
RA Isogai T., Ota T., Hayashi K., Sugiyama T., Otsuki T., Suzuki Y.,  
RA Nishikawa T., Nagai K., Sugano S., Shiratori A., Sudo H., Sugawara M.,  
RA Nagatsuma M., Hosoi T., Kaku Y., Kodaira H., Kondo H., Sugawara M.,  
RA Takahashi M., Chiba Y., Ishida S., Murakawa K., Ono Y., Takiguchi S.,  
RA Watanabe S., Kimura K., Murakami K., Ishii S., Kawai Y., Saito K.,  
RA Yamamoto J., Wakamatsu A., Nakamura Y., Nagahari K., Masuho Y.,  
RA Ninomiya K., Iwayanagi T.;  
RT "NEDO human cDNA sequencing project.";  
RL Submitted (MAY-2001) to the EMBL/GenBank/DBJ databases.  
DR EMBL; AK027648; BAB5284.1; -.  
DR GenBank; HGNC:19316; LEPREL.  
DR GO; GO:0016706; F:oxidoreductase activity, acting on paired d. . . ; IEA.  
DR GO; GO:0019538; P:protein metabolism; IEA.  
DR InterPro; IPR005123; 2OG-Fell\_Oxy.  
DR InterPro; IPR006620; Pro\_4\_hyd\_alph.  
DR InterPro; IPR008941; TPR-like.  
DR Pfam; PF03171; 2OG-Fell\_Oxy; 1.  
DR SMART; SMO0702; P4HC; 1.  
DR Hypothetical protein.  
KW SEQUENCE 804 AA; 90544 MW; 23C99DE603D208CD CRC64;  
Query Match 38.2%; Score 281; DB 4; Length 804;  
Best Local Similarity 99.4%; Pred. No. 2.4e-277;  
Matches 681; Conservative 0; Mismatches 4; Indels 0; Gaps 0;  
QY 1 MAVRALKLLTLLAVAAASQAEVSEAGMGWTPDLLFAEGTAAAYARGDWPVGLSMER 60  
DB 1 MAVRALKLLTLLAVAAASQAEVSEAGMGWTPDLLFAEGTAAAYARGDWPVGLSMER 60  
QY 61 ALRSHAALRALRLCRTOCAADFPWELDPDPSPAQASGAGALRDLSPFGGLLRRACL 120  
DB 61 ALRSHAALRALRLCRTOCAADFPWELDPDPSPAQASGAGALRDLSPFGGLLRRACL 120  
QY 121 RRLGLPPAAHSLSEMELEFRKSPYNYLQVAYFKINKLEKAVAAHTFFVGNPEHMEMQ 180  
DB 121 RRLGLPPAAHSLSEMELEFRKSPYNYLQVAYFKINKLEKAVAAHTFFVGNPEHMEMQ 180  
QY 181 QNLDDYQTMGSKVKAADFKDLETQPHMQEFLGVRLYSEEQPQEAHPHLEALQYFVAYE 240  
DB 181 QNLDDYQTMGSKVKAADFKDLETQPHMQEFLGVRLYSEEQPQEAHPHLEALQYFVAYE 240  
QY 241 ECRALCEGPDYDGVNYLEYNADLFOAITDHYIOVLNCKQNCVTELASHPSREKPFDFL 300  
DB 241 ECRALCEGPDYDGVNYLEYNADLFOAITDHYIOVLNCKQNCVTELASHPSREKPFDFL 300  
QY 301 PSHYNLOFAYYNYIGNYTOAGCAKTYLLFPFNDVWQNLAYYAAAMLGEHSTRSIGPRE 360  
DB 301 PSHYNLOFAYYNYIGNYTOAGCAKTYLLFPFNDVWQNLAYYAAAMLGEHSTRSIGPRE 360  
QY 361 SAKERYQSLLLEKELFFAYDFVGFIPFVDPDSWTPPEVVKRLQEKSERETAVRISQE 420  
DB 361 SAKERYQSLLLEKELFFAYDFVGFIPFVDPDSWTPPEVVKRLQEKSERETAVRISQE 420  
QY 421 IGNLMKEITLVBEKTESLDVSRITREGGPIIYEGISLTWNSKLLNGYQRVMDGVISD 480  
DB 421 IGNLMKEITLVBEKTESLDVSRITREGGPIIYEGISLTWNSKLLNGYQRVMDGVISD 480  
QY 481 HECQLRLTNVAATSGDGYRGOTSPHTNEKPYGVTVFKALKLGOEKVPLQSAHLYNN 540  
DB 481 HECQLRLTNVAATSGDGYRGOTSPHTNEKPYGVTVFKALKLGOEKVPLQSAHLYNN 540  
QY 541 VTEKVRIMESYFRLDTPLYFSYSHLVCRTAIEVQAEKDDSHPVHVDNCILNAETLVC 600  
DB 541 VTEKVRIMESYFRLDTPLYFSYSHLVCRTAIEVQAEKDDSHPVHVDNCILNAETLVC 600  
QY 601 VKEPPAYTFRDYSAIYLNGDFDGGNFYFTELDKTVTAEVQPCGQRAVGFSSGTENPHG 660  
DB 601 VKEPPAYTFRDYSAIYLNGDFDGGNFYFTELDKTVTAEVQPCGQRAVGFSSGTENPHG 660

```
QY 661 VKAVTRGQRCALWFTLDRHSR 685
Db 661 VKAVTRGQRCALWFTLDRHSR 685

RESULT 7
Q9H6K3 PRELIMINARY; PRT; 390 AA.
AC Q9H6K3
DT 01-MAR-2001 (TReMBLrel. 16, Created)
DT 01-MAR-2001 (TReMBLrel. 16, Last sequence update)
DT 01-OCT-2003 (TReMBLrel. 25, Last annotation update)
DE Hypothetical protein FLJ22188.
OS Homo sapiens (human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
ON NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RA Watanabe K., Kumagai A., Itakura S., Yamazaki M., Tashiro H., Ota T.,
RA Suzuki Y., Obayashi M., Nishi T., Shibahara T., Tanaka T.,
RA Nakamura Y., Isogai T., Sugano S.;
RT "NEDO human cDNA sequencing project.";
RL Submitted (AUG-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL; AK025841; BAB15256.1; -.
DR GO; GO:0016706; P:oxidoreductase activity, acting on paired d. . .; IEA.
DR GO; GO:0019538; P:protein metabolism; IEA.
DR InterPro; IPR005123; ZOG-FcII_Oxy.
DR InterPro; IPR00886; ER_target_S.
DR InterPro; IPR006620; Pro_4_hyd_alph.
DR Pfam; PF03171; ZOG-FcII_Oxy; 1.
DR SMART; SM00702; P4HC; 1.
DR PROSITE; PS00014; ER_TARGET; 1.
DR KW Hypothetical protein.
SQ SEQUENCE 390 AA; 44093 MW; 61D2999310B677FB CRC64;

Query Match 25.5%; Score 188; DB 4; Length 390;
Best Local Similarity 99.5%; Pred. No. 1.1e-182;
Matches 388; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 347 MLGSEHTRSGPRESAKEYQSRILKEKLLFFAYDVGIPFVDPDSWTPPEVTPKRLQEK 406
Db 1 MLGSEHTRSGPRESAKEYQSRILKEKLLFFAYDVGIPFVDPDSWTPPEVTPKRLQEK 60

QY 407 QKSERETAVRISOEIGNLMKEIETLVEEKTKEISLDVSRITREGGPLYEGISLTWNSKLL 466
Db 61 QKSERETAVRISOEIGNLMKEIETLVEEKTKEISLDVSRITREGGPLYEGISLTWNSKLL 120

QY 467 NGQVRVMDGVI SDHECQELQRLTNVAATSGDGYRGQTSPTNEKFGYGVTVFKALKLQ 526
Db 121 NGQVRVMDGVI SDHECQELQRLTNVAATSGDGYRGQTSPTNEKFGYGVTVFKALKLQ 180

QY 527 EGKVPQSAHLNVYVTEKVRIMESYFRDLTPLYFSYSHLVCRITAEEVQAEKDDSHPV 586
Db 181 EGKVPQSAHLNVYVTEKVRIMESYFRDLTPLYFSYSHLVCRITAEEVQAEKDDSHPV 240

QY 587 HVNDCILNAETLVCKEPPAYTRDYSAILYLNGDFDGGNFYFTELDKAKTVAEVOPOCG 646
Db 241 HVNDCILNAETLVCKEPPAYTRDYSAILYLNGDFDGGNFYFTELDKAKTVAEVOPOCG 300

QY 647 RAVGFSSGTENPHGVKAVTGQRCALWFTLDRHSRDRVQADDLVKMLFSPEDMDS 706
Db 301 RAVGFSSGTENPHGVKAVTGQRCALWFTLDRHSRDRVQADDLVKMLFSPEDMDS 360

QY 707 QEQLDAQQPPPEPAQESLSGSESKPKDEL 736
Db 361 QEQLDAQQPPPEPAQESLSGSESKPKDEL 390

RESULT 8
Q9RIJ8 PRELIMINARY; PRT; 728 AA.
ID Q9RIJ8
```

```
Q9RIJ8;
01-MAY-2000 (TReMBLrel. 13, Created)
01-MAY-2000 (TReMBLrel. 13, Last sequence update)
01-OCT-2003 (TReMBLrel. 25, Last annotation update)
DE Lepreacan.
GN LEPRE1.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
ON NCBI_TaxID=10116;
RN [1]
RP SEQUENCE FROM N.A.
RA Wassenhove-McCarthy D.J., McCarthy K.J.;
RT "Molecular characterization of a novel basement membrane-associated
RT proteoglycan, leprecan.";
RL J. Biol. Chem. 274:25004-25017(1999).
DR EMBL; AF087433; AAD51875.1; -.
DR GO; GO:0016706; P:oxidoreductase activity, acting on paired d. . .; IEA.
DR GO; GO:0019538; P:protein metabolism; IEA.
DR InterPro; IPR005123; ZOG-FcII_Oxy.
DR InterPro; IPR00886; ER_target_S.
DR InterPro; IPR006620; Pro_4_hyd_alph.
DR InterPro; IPR008941; TPR-like.
DR Pfam; PF03171; ZOG-FcII_Oxy; 1.
DR SMART; SM00702; P4HC; 1.
DR PROSITE; PS00014; ER_TARGET; 1.
DR KW Hypothetical protein.
SQ SEQUENCE 728 AA; 82389 MW; 06AFE6972BF3EE1F CRC64;

Query Match 13.7%; Score 101; DB 11; Length 728;
Best Local Similarity 100.0%; Pred. No. 1.3e-93;
Matches 101; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 602 KEPPAYTRDYSAILYLNGDFDGGNFYFTELDKAKTVAEVOPOCGRAVGFSSGTENPHGV 661
Db 594 KEPPAYTRDYSAILYLNGDFDGGNFYFTELDKAKTVAEVOPOCGRAVGFSSGTENPHGV 653

QY 662 KAVTRGQRCALWFTLDRHSRDRVQADDLVKMLFSPPEE 702
Db 654 KAVTRGQRCALWFTLDRHSRDRVQADDLVKMLFSPPEE 694

RESULT 9
Q8CFL3 PRELIMINARY; PRT; 739 AA.
ID Q8CFL3
AC Q8CFL3;
DT 01-MAR-2003 (TReMBLrel. 23, Created)
DT 01-MAR-2003 (TReMBLrel. 23, Last sequence update)
DT 01-OCT-2003 (TReMBLrel. 25, Last annotation update)
DE Similar to leprecan.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.
ON NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J; TISSUE=Breast tumor;
RA Strausberg R.;
RL Submitted (FEB-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL; BC024047; AAH24047.1; -.
DR PIR; PTO551; PTO551.
DR GO; GO:0016706; P:oxidoreductase activity, acting on paired d. . .; IEA.
DR InterPro; IPR005123; ZOG-FcII_Oxy.
DR InterPro; IPR00886; ER_target_S.
DR InterPro; IPR006620; Pro_4_hyd_alph.
DR InterPro; IPR008941; TPR-like.
DR Pfam; PF03171; ZOG-FcII_Oxy; 1.
DR SMART; SM00702; P4HC; 1.
DR PROSITE; PS00014; ER_TARGET; 1.
DR KW Hypothetical protein.
SQ SEQUENCE 739 AA; 83592 MW; 64B7A568B80B68EA CRC64;

Query Match 13.7%; Score 101; DB 11; Length 739;
```

```

Best Local Similarity 100.0%; Pred. No. 1.3e-93;
Matches 101; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 602 KPPAYTFRDYSAILYLNGDFDGNFYFTLDAKTVAEVPQCGRAVFGSSGTENPHGV 661
DB 605 KPPAYTFRDYSAILYLNGDFDGNFYFTLDAKTVAEVPQCGRAVFGSSGTENPHGV 664

QY 662 KAVTRGQRCATLWFTLDPHRSERDRVQADDLVKMLFSPPE 702
DB 665 KAVTRGQRCATLWFTLDPHRSERDRVQADDLVKMLFSPPE 705

RESULT 10
Q9CWK5 PRELIMINARY; PRT; 742 AA.
AC Q9CWK5;
DT 01-JUN-2001 (TRENBLrel. 17, Created)
DT 01-JUN-2001 (TRENBLrel. 17, Last sequence update)
DT 01-OCT-2003 (TRENBLrel. 25, Last annotation update)
DE 2410024C15RIK protein.
GN LEPRE1 OR GROS1 OR 2410024C15RIK.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=CD1 ICR; TISSUE=Embryonic stem cells;
RX MEDLINE=21085660; PubMed=1217851;
RA Kawai J., Shinagawa A., Shibata K., Yoshino M., Itoh M., Ishii Y.,
RA Aikawa T., Hara A., Fukunishi Y., Konno H., Adachi J., Fukuda S.,
RA Aizawa K., Izawa M., Nishi K., Kiyosawa H., Kondo S., Yamana K.,
RA Saito T., Okazaki Y., Gojohori T., Bono H., Kasukawa T., Saito R.,
RA Kadota K., Matsuda H.A., Ashburner M., Batalov S., Casavant T.,
RA Fleischmann W., Gaasterland T., Gissi C., King B., Kochiwa H.,
RA Kuehl P., Lewis S., Matsuo Y., Nikaido I., Pesole G., Quackenbush J.,
RA Schriml L.M., Straubli F., Suzuki R., Tomita M., Wagner L., Washio T.,
RA Sakai K., Okido T., Furuno M., Aono H., Baldarelli R., Barsh G.,
RA Blake J., Boffelli D., Bojunga N., Carninci P., de Bonaldi M.F.,
RA Brownstein M.J., Bult C., Fletcher C., Fujita M., Gariboldi M.,
RA Gusticich S., Hill D., Hofmann M., Hume D.A., Kamiya M., Lee N.H.,
RA Lyons P., Marchionni L., Mashima J., Mazzarelli J., Mombaerts P.,
RA Nordone P., Ring B., Ringwald M., Rodriguez I., Sakamoto N.,
RA Sasaki H., Sato K., Schoenbach C., Seya T., Shibata Y., Storch K.-F.,
RA Suzuki H., Toyooka K., Wang K.H., Weitz C., Whittaker C., Wilming L.,
RA Wynshaw-Boris A., Yoshida K., Hasegawa Y., Kawaji H., Kohtsuki S.,
RA Hayashizaki Y.;
RT "Functional annotation of a full-length mouse cDNA collection.";
RL Nature 409:685-690(2001).
DR EMBL; AK010578; BAB27041.1; -.
DR MGD; MGI:1888921; Leprel.
DR GO; GO:0005737; C:cytoplasm; IDA.
DR GO; GO:0005634; C:nucleus; IDA.
DR GO; GO:0016049; P:cell growth; IDA.
DR InterPro; IPR005123; 2OG-Fell_Oxy.
DR InterPro; IPR006620; Pro 4 hvd_alph.
DR InterPro; IPR008941; TPR-like.
DR Pfam; PF03171; 2OG-Fell_Oxy; 1.
DR SMART; SM00702; P4HC; 1.
SQ SEQUENCE 742 AA; 84086 MW; CFC44D34D96ACEA7 CRC64;

Query Match 13.7%; Score 101; DB 11; Length 742;
Best Local Similarity 100.0%; Pred. No. 1.3e-93;
Matches 101; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 602 KPPAYTFRDYSAILYLNGDFDGNFYFTLDAKTVAEVPQCGRAVFGSSGTENPHGV 661
DB 605 KPPAYTFRDYSAILYLNGDFDGNFYFTLDAKTVAEVPQCGRAVFGSSGTENPHGV 664

QY 662 KAVTRGQRCATLWFTLDPHRSERDRVQADDLVKMLFSPPE 702
DB 665 KAVTRGQRCATLWFTLDPHRSERDRVQADDLVKMLFSPPE 705

RESULT 11
Q9QZT7 PRELIMINARY; PRT; 747 AA.
AC Q9QZT7;
DT 01-MAY-2000 (TRENBLrel. 13, Created)
DT 01-MAY-2000 (TRENBLrel. 13, Last sequence update)
DT 01-OCT-2003 (TRENBLrel. 25, Last annotation update)
DE Growth suppressor 1L.
GN LEPRE1 OR GROS1.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=CD1 ICR; TISSUE=Fibroblast;
RX MEDLINE=20406537; PubMed=10951563;
RA Kaul S.C., Sugihara T., Yoshida A., Nomura H., Wadhwa R.;
RT "Gros1, a potential growth suppressor on chromosome 1: its identity to
RT basement membrane-associated proteoglycan, leprecan.";
RL Oncogene 19:3576-3583(2000).
DR EMBL; AF165163; AAF04806.1; -.
DR MGD; MGI:1888921; Leprel.
DR GO; GO:0005737; C:cytoplasm; IDA.
DR GO; GO:0005634; C:nucleus; IDA.
DR GO; GO:0016049; P:cell growth; IDA.
DR InterPro; IPR005123; 2OG-Fell_Oxy.
DR InterPro; IPR006620; Pro 4 hvd_alph.
DR InterPro; IPR008941; TPR-like.
DR Pfam; PF03171; 2OG-Fell_Oxy; 1.
DR SMART; SM00702; P4HC; 1.
SQ SEQUENCE 747 AA; 84811 MW; EBEAB2A62E824123 CRC64;

Query Match 12.4%; Score 91; DB 11; Length 747;
Best Local Similarity 100.0%; Pred. No. 2.1e-83;
Matches 91; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 612 YSAILYLNGDFDGNFYFTLDAKTVAEVPQCGRAVFGSSGTENPHGVKAVTRGQRC 671
DB 614 YSAILYLNGDFDGNFYFTLDAKTVAEVPQCGRAVFGSSGTENPHGVKAVTRGQRC 673

QY 672 IALWFTLDPHRSERDRVQADDLVKMLFSPPE 702
DB 674 IALWFTLDPHRSERDRVQADDLVKMLFSPPE 704

RESULT 12
Q9QZT6 PRELIMINARY; PRT; 542 AA.
AC Q9QZT6;
DT 01-MAY-2000 (TRENBLrel. 13, Created)
DT 01-MAY-2000 (TRENBLrel. 13, Last sequence update)
DT 01-OCT-2003 (TRENBLrel. 25, Last annotation update)
DE Growth suppressor 1S.
GN LEPRE1 OR GROS1.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=CD1 ICR;
RX MEDLINE=20406537; PubMed=10951563;
RA Kaul S.C., Sugihara T., Yoshida A., Nomura H., Wadhwa R.;
RT "Gros1, a potential growth suppressor on chromosome 1: its identity to
RT basement membrane-associated proteoglycan, leprecan.";
RL Oncogene 19:3576-3583(2000).
DR EMBL; AF165164; AAF04807.1; -.
DR MGD; MGI:1888921; Leprel.
DR GO; GO:0005737; C:cytoplasm; IDA.
DR GO; GO:0005634; C:nucleus; IDA.
DR GO; GO:0016049; P:cell growth; IDA.

```

```
DR InterPro; IPR008941; TPR-like.
SQ SEQUENCE 542 AA; 61599 MW; 0A211411F71DF16B CRC64;

Query Match
Best Local Similarity 11.1%; Score 82; DB 11; Length 542;
Matches 82; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 383 FGIPFVDPDSWTPEEVIKPELQKQKSERETAVRISQEIQLNLMKETETLVESKTKESLDV 442
DB 385 FGIPFVDPDSWTPEEVIKPELQKQKSERETAVRISQEIQLNLMKETETLVESKTKESLDV 444
QY 443 SRLTREGGPLLYEGISLTMNSK 464
DB 445 SRLTREGGPLLYEGISLTMNSK 466

RESULT 13
Q8BSV2 Q8BSV2 PRELIMINARY; PRT; 738 AA.
AC Q8BSV2;
DT 01-MAR-2003 (T-EMBLrel. 23, Created)
DT 01-MAR-2003 (T-EMBLrel. 23, Last sequence update)
DT 01-OCT-2003 (T-EMBLrel. 25, Last annotation update)
DE Leprecan 1.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J; TISSUE=pituitary;
RX MEDLINE=22354683; PubMed=12466851;
RA The FANTOM Consortium.
RT "Analysis of the mouse transcriptome based on functional annotation of
RT 60,770 full-length cDNAs."
RL Nature 420:563-573(2002).
DR ENBL; AK030436; BAC26962.1; -.
DR MGI; AK076434; BAC36342.1; -.
DR GO; GO:0016706; P:oxidoreductase activity, acting on paired d. . .; IEA.
DR GO; GO:0019538; P:protein metabolism; IEA.
DR InterPro; IPR005123; 2OG-Fell_Oxy.
DR InterPro; IPR000886; ER_target_S.
DR InterPro; IPR006620; Pro_4_hyd_alph.
DR InterPro; IPR008941; TPR-like.
DR Pfam; PF03171; 2OG-Fell_Oxy; 1.
DR SMART; SM00702; P4HC; 1.
DR PROSITE; PS00014; ER_TARGET; 1.
KW Hypothetical protein.
FT NON_TER 1
SQ SEQUENCE 409 AA; 46325 MW; 49CB7CAA66F909FC CRC64;

Query Match
Best Local Similarity 2.2%; Score 16; DB 11; Length 409;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 656 ENPHGVKAVTRGQRC 671
DB 350 ENPHGVKAVTRGQRC 365

RESULT 15
Q8CG71 Q8CG71 PRELIMINARY; PRT; 703 AA.
AC Q8CG71;
DT 01-MAR-2003 (T-EMBLrel. 23, Created)
DT 01-MAR-2003 (T-EMBLrel. 23, Last sequence update)
DT 01-OCT-2003 (T-EMBLrel. 25, Last annotation update)
DE Myxoid liposarcoma associated protein 4.
GN AW553532 OR MLAT4.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL;
RA Thein-Jarnum S., Aman P.;
RT "The novel gene Myxoid Liposarcoma Associated Transcript 4 (MLAT4)
RT belongs to the Grosi/leprecan proteoglycan family."
RL Submitted (FEB-2002) to the EMBL/GenBank/DBJ databases.
DR ENBL; AJ430350; CAD23038.1; -.
DR MGI; MGI:2146663; AW553532.
DR GO; GO:0016706; P:oxidoreductase activity, acting on paired d. . .; IEA.
DR GO; GO:0019538; P:protein metabolism; IEA.
DR InterPro; IPR005123; 2OG-Fell_Oxy.
DR InterPro; IPR000886; ER_target_S.
DR InterPro; IPR006620; Pro_4_hyd_alph.
DR InterPro; IPR001440; TPR.
DR InterPro; IPR008941; TPR-like.
DR Pfam; PF03171; 2OG-Fell_Oxy; 1.
DR SMART; SM00702; P4HC; 1.
DR PROSITE; PS00014; ER_TARGET; 1.
SQ SEQUENCE 703 AA; 80154 MW; 1F42F9B9938573E4 CRC64;

Query Match
Best Local Similarity 2.2%; Score 16; DB 11; Length 703;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
DR InterPro; IPR008941; TPR-like.
SQ SEQUENCE 542 AA; 61599 MW; 0A211411F71DF16B CRC64;

Query Match
Best Local Similarity 11.1%; Score 82; DB 11; Length 542;
Matches 82; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 383 FGIPFVDPDSWTPEEVIKPELQKQKSERETAVRISQEIQLNLMKETETLVESKTKESLDV 442
DB 385 FGIPFVDPDSWTPEEVIKPELQKQKSERETAVRISQEIQLNLMKETETLVESKTKESLDV 444
QY 443 SRLTREGGPLLYEGISLTMNSK 464
DB 445 SRLTREGGPLLYEGISLTMNSK 466

RESULT 13
Q8BSV2 Q8BSV2 PRELIMINARY; PRT; 738 AA.
AC Q8BSV2;
DT 01-MAR-2003 (T-EMBLrel. 23, Created)
DT 01-MAR-2003 (T-EMBLrel. 23, Last sequence update)
DT 01-OCT-2003 (T-EMBLrel. 25, Last annotation update)
DE Leprecan 1.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J; TISSUE=pituitary;
RX MEDLINE=22354683; PubMed=12466851;
RA The FANTOM Consortium.
RT "Analysis of the mouse transcriptome based on functional annotation of
RT 60,770 full-length cDNAs."
RL Nature 420:563-573(2002).
DR ENBL; AK030436; BAC26962.1; -.
DR MGI; AK076434; BAC36342.1; -.
DR GO; GO:0016706; P:oxidoreductase activity, acting on paired d. . .; IEA.
DR GO; GO:0019538; P:protein metabolism; IEA.
DR InterPro; IPR005123; 2OG-Fell_Oxy.
DR InterPro; IPR000886; ER_target_S.
DR InterPro; IPR006620; Pro_4_hyd_alph.
DR InterPro; IPR008941; TPR-like.
DR Pfam; PF03171; 2OG-Fell_Oxy; 1.
DR SMART; SM00702; P4HC; 1.
DR PROSITE; PS00014; ER_TARGET; 1.
SQ SEQUENCE 738 AA; 83528 MW; 75967DE318F55F4A CRC64;

Query Match
Best Local Similarity 10.9%; Score 80; DB 11; Length 738;
Matches 80; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 602 KEPPAYTRDYSAIYLNGDFDGNFYFTLDAKTVAEYVQPCGQRAVGFSSGTENPHGV 661
DB 604 KEPPAYTRDYSAIYLNGDFDGNFYFTLDAKTVAEYVQPCGQRAVGFSSGTENPHGV 663
QY 662 KAVTRGQRCALWFTLDP 681
DB 664 KAVTRGQRCALWFTLDP 683

RESULT 14
Q8C673 Q8C673 PRELIMINARY; PRT; 409 AA.
AC Q8C673;
DT 01-MAR-2003 (T-EMBLrel. 23, Created)
DT 01-MAR-2003 (T-EMBLrel. 23, Last sequence update)
DT 01-OCT-2003 (T-EMBLrel. 25, Last annotation update)
DE Hypothetical protein (Fragment).
GN AW553532.
OS Mus musculus (Mouse).
```

Oy 656 ENPHGVKAVTRGQCA 571  
| | | | | | | | | |  
Db 644 ENPHGVKAVTRGQCA 659

Search completed: July 18, 2004, 09:59:46  
Job time : 54 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - nucleic search, using frame\_plus\_p2n model

Run on: July 18, 2004, 18:12:26 ; Search time 148 Seconds  
(without alignments)  
2759.756 Million cell updates/sec

Title: US-10-045-815-4

Perfect score: 736

Sequence: 1 MAVRALKLLTLLAVVAAS.....PPEPAQESUSGSSEKPKDEL 736

Scoring table:

OLIGO  
Xgapop 60.0 , Xgapext 60.0  
Ygapop 60.0 , Ygapext 60.0  
Fgapop 6.0 , Fgapext 7.0  
Delop 6.0 , Delext 7.0

Searched: 682709 seqs, 277475446 residues

Word size: 1

Total number of hits satisfying chosen parameters: 1360453

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Command line parameters:

-MODEL=frame+ p2n.model -DEV=xlp  
-Q=/cgn2\_1/USPTO.spool\_P/US10045815/runat\_14072004\_123119\_17267/app\_query.fasta\_1.903  
-DB=Issued Patents NA -GFFT=fastap -SUFFIX=oli.rni -MINMATCH=0.1 -LOOPCL=0  
-LOOPEXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=oligo -TRANS=human40.cdi  
-LIST=45 -DOCALIGN=200 -THR SCORE=quality -THR MIN=1 -ALIGN=15 -MODE=LOCAL  
-OUTFMT=ptco -NORM=ext -HEADSIZE=500 -MINLEN=0 -MAXLEN=200000000  
-USR=US10045815@cgn\_1\_128@runat\_14072004\_123119\_17267 -NCPU=6 -ICPU=3  
-NO MMAP -LARGEQUERY -NEG SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG  
-DEV TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=60 -XGAPEXT=60 -FGAPOP=6  
-FGAPEXT=7 -YGAPOP=60 -YGAPEXT=60 -DELOP=6 -DELEXT=7

Database :

Issued Patents NA.\*  
1: /cgn2\_6/ptodata/2/ina/5A COMB.seq.\*  
2: /cgn2\_6/ptodata/2/ina/5B COMB.seq.\*  
3: /cgn2\_6/ptodata/2/ina/6A COMB.seq.\*  
4: /cgn2\_6/ptodata/2/ina/6B COMB.seq.\*  
5: /cgn2\_6/ptodata/2/ina/PTCUS COMB.seq.\*  
6: /cgn2\_6/ptodata/2/ina/backfiles1.seq.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
C 1	9	1.2	13856	4	US-08-956-171E-31
C 2	9	1.2	37769	4	US-08-311-731A-23
C 3	9	1.2	4403765	3	US-09-103-840A-2
C 4	9	1.2	4411529	3	US-09-103-840A-1
C 5	8	1.1	251	4	US-08-956-171E-4554
C 6	8	1.1	301	4	US-09-439-313-271
C 7	8	1.1	301	4	US-09-352-616A-271
C 8	8	1.1	301	4	US-09-232-149A-271
C 9	8	1.1	301	4	US-09-159-812-271
C 10	8	1.1	301	4	US-09-636-215-271
C 11	8	1.1	301	4	US-09-685-166A-271
C 12	8	1.1	301	4	US-09-688-489-271

C 13	8	1.1	363	4	US-08-407-620A-12	Sequence 12, Appl
C 14	8	1.1	375	4	US-09-252-991A-4191	Sequence 4191, Ap
C 15	8	1.1	387	4	US-09-582-337-21	Sequence 21, Appl
C 16	8	1.1	408	4	US-09-075-460-14	Sequence 14, Appl
C 17	8	1.1	480	3	US-09-039-555B-11	Sequence 11, Appl
C 18	8	1.1	646	4	US-09-614-912-135	Sequence 135, App
C 19	8	1.1	657	4	US-09-252-991A-4339	Sequence 4339, Ap
C 20	8	1.1	663	4	US-09-252-991A-3252	Sequence 3252, Ap
C 21	8	1.1	789	4	US-09-252-991A-4083	Sequence 4083, Ap
C 22	8	1.1	855	4	US-09-252-991A-4107	Sequence 4107, Ap
C 23	8	1.1	894	4	US-09-328-352-1260	Sequence 1260, Ap
C 24	8	1.1	987	4	US-09-221-017B-1115	Sequence 1115, Ap
C 25	8	1.1	1105	4	US-09-221-017B-108	Sequence 108, App
C 26	8	1.1	1155	4	US-09-252-991A-4515	Sequence 4515, Ap
C 27	8	1.1	1281	4	US-09-252-991A-3691	Sequence 3691, Ap
C 28	8	1.1	1515	4	US-09-252-991A-4123	Sequence 4123, Ap
C 29	8	1.1	1602	4	US-09-252-991A-3776	Sequence 3776, Ap
C 30	8	1.1	1608	2	US-08-211-718-8	Sequence 8, Appli
C 31	8	1.1	1755	4	US-09-252-991A-3588	Sequence 3588, Ap
C 32	8	1.1	1842	4	US-09-489-039A-2888	Sequence 2888, Ap
C 33	8	1.1	2004	1	US-08-471-033-18	Sequence 18, Appl
C 34	8	1.1	2004	2	US-08-471-044-18	Sequence 18, Appl
C 35	8	1.1	2004	2	US-08-463-483A-18	Sequence 18, Appl
C 36	8	1.1	2004	2	US-08-471-046A-18	Sequence 18, Appl
C 37	8	1.1	2004	2	US-08-470-566B-18	Sequence 18, Appl
C 38	8	1.1	2004	2	US-08-469-334-18	Sequence 18, Appl
C 39	8	1.1	2004	3	US-09-300-529-18	Sequence 18, Appl
C 40	8	1.1	2576	1	US-08-471-033-35	Sequence 35, Appl
C 41	8	1.1	2576	2	US-08-471-044-35	Sequence 35, Appl
C 42	8	1.1	2576	2	US-08-463-483A-35	Sequence 35, Appl
C 43	8	1.1	2576	2	US-08-471-046A-35	Sequence 35, Appl
C 44	8	1.1	2576	2	US-08-470-566B-35	Sequence 35, Appl
C 45	8	1.1	2576	2	US-08-469-334-35	Sequence 35, Appl

ALIGNMENTS

RESULT 1

US-08-956-171E-31/c  
; Sequence 31, Application US/08956171E  
; Patent No. 6593114

GENERAL INFORMATION:  
APPLICANT: Charles Kunsch

Gil H. Choi  
Patrick S. Dillon  
Craig A. Rosen  
Steven C. Barash

Michael R. Fannon  
TITLE OF INVENTION: Staphylococcus aureus Polynucleotides and Sequences

NUMBER OF SEQUENCES: 5256

CORRESPONDENCE ADDRESS:

ADDRESSEE: Human Genome Sciences, Inc.  
STREET: 9410 Key West Avenue  
CITY: Rockville

STATE: Maryland  
COUNTRY: USA  
ZIP: 20850

COMPUTER READABLE FORM:

MEDIUM TYPE: Diskette, 3.50 inch, 1.4Mb storage

COMPUTER: HP Vectra 486/33

OPERATING SYSTEM: MSDOS version 6.2

SOFTWARE: ASCII Text

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/956,171E

FILING DATE: 20-Oct-1997

CLASSIFICATION: <Unknown>

PRIOR APPLICATION DATA:

APPLICATION NUMBER: 60/009,861

FILING DATE: January 5, 1996

APPLICATION NUMBER: 08/781,986

FILING DATE: January 3, 1997

ATTORNEY/AGENT INFORMATION:

NAME: Mark J. Hyman



REGISTRATION NUMBER: 46,789  
REFERENCE/DOCKET NUMBER: PB248P1  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (240) 314-1224  
TELEFAX: (301) 309-8439  
INFORMATION FOR SEQ ID NO: 31:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 13856 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: double  
TOPOLOGY: linear  
SEQUENCE DESCRIPTION: SEQ ID NO: 31:  
US-08-956-171E-31

Alignment Scores: 183 Length: 13856  
Pred. No.: 9.00 Matches: 9  
Score: 100.00% Conservative: 0  
Best Local Similarity: 100.00% Mismatches: 0  
Query Match: 1.22% Indels: 0  
DB: 4 Gaps: 0

US-10-045-815-4 (1-736) x US-08-956-171E-31 (1-13856)

Qy 557 ThrProLeuTyrPheSerTyrSerHis 565  
Db 5923 ACTCCTTAATATTTCTTACTACAT 5897

## RESULT 2

US-08-311-731A-23/c

; Sequence 23, Application US/08311731A

; Patent No. 6583266

; GENERAL INFORMATION:

; APPLICANT: SMITH, DOUGLAS

; APPLICANT: MAO, JEN-I

; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES

; TITLE OF INVENTION: RELATING TO MYCOBACTERIUM TUBERCULOSIS AND LAPRAE FOR

; TITLE OF INVENTION: DIAGNOSTICS AND THERAPEUTICS

; NUMBER OF SEQUENCES: 411

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: WOLF, GREENFIELD & SACKS, P.C.

; STREET: 600 ATLANTIC AVENUE

; CITY: BOSTON

; STATE: MASSACHUSETTS

; COUNTRY: USA

; ZIP: 02210

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Floppy disk

; COMPUTER: IBM PC compatible

; OPERATING SYSTEM: PC-DOS/MS-DOS

; SOFTWARE: PatentIn Release #1.0, Version #1.25

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/08/311.731A

; FILING DATE:

; CLASSIFICATION: 530

; ATTORNEY/AGENT INFORMATION:

; NAME: GATES, EDWARD R.

; REGISTRATION NUMBER: 31,616

; REFERENCE/DOCKET NUMBER: C0044/7125

; TELECOMMUNICATION INFORMATION:

; TELEPHONE: 617/720-3500

; TELEFAX: 617/720-2441

; INFORMATION FOR SEQ ID NO: 23:

; SEQUENCE CHARACTERISTICS:

; LENGTH: 37769 base pairs

; TYPE: nucleic acid

; STRANDEDNESS: double

; TOPOLOGY: circular

; MOLECULE TYPE: DNA (genomic)

; HYPOTHETICAL: NO

; ANTI-SENSE: NO

; ORIGINAL SOURCE:

; ORGANISM: MYCOBACTERIUM LAPRAE

US-08-311-731A-23

Alignment Scores:

Pred. No.: 483 Length: 37769  
Score: 9.00 Matches: 9  
Percent Similarity: 100.00% Conservative: 0  
Best Local Similarity: 100.00% Mismatches: 0  
Query Match: 1.22% Indels: 0  
DB: 4 Gaps: 0

US-10-045-815-4 (1-736) x US-08-311-731A-23 (1-37769)

Qy 646 GlyArgAlaValGlyPheSerSerGly 654

Db 28383 GGTGAGCGGCGCGGTTTTCGTCCGG 28357

## RESULT 3

US-09-103-840A-2

; Sequence 2, Application US/09103840A

; Patent No. 6294328

; GENERAL INFORMATION:

; APPLICANT: FLEISCHMAN, Robert D.

; APPLICANT: WHITE, Owen R.

; APPLICANT: FRASER, Claire M.

; APPLICANT: VENTER, John C.

; TITLE OF INVENTION: DNA SEQUENCES FOR STRAIN ANALYSIS IN MYCOBACTERIUM

; TITLE OF INVENTION: TUBERCULOSIS

; FILE REFERENCE: 24366-20007.00

; CURRENT APPLICATION NUMBER: US/09/103.840A

; CURRENT FILING DATE: 1998-06-24

; NUMBER OF SEQ ID NOS: 2

; SOFTWARE: PatentIn Ver. 2.1

; SEQ ID NO 2

; LENGTH: 4403765

; TYPE: DNA

; ORGANISM: Mycobacterium tuberculosis

; FEATURE:

; OTHER INFORMATION: CDC 1551

; OTHER INFORMATION: "n" bases at various positions throughout the sequence

; OTHER INFORMATION: represent a, t, c or g

US-09-103-840A-2

Alignment Scores:

Pred. No.: 4.68e+04 Length: 4403765  
Score: 9.00 Matches: 9  
Percent Similarity: 100.00% Conservative: 0  
Best Local Similarity: 100.00% Mismatches: 0  
Query Match: 1.22% Indels: 0  
DB: 3 Gaps: 0

US-10-045-815-4 (1-736) x US-09-103-840A-2 (1-4403765)

Qy 14 AlaValAlaAlaAlaSerGlnAla 22

Db 3088113 GCCGTCGCGGCGCGGCGGCGGCG 3088139

## RESULT 4

US-09-103-840A-1

; Sequence 1, Application US/09103840A

; Patent No. 6294328

; GENERAL INFORMATION:

; APPLICANT: FLEISCHMAN, Robert D.

; APPLICANT: WHITE, Owen R.

; APPLICANT: FRASER, Claire M.

; APPLICANT: VENTER, John C.

; TITLE OF INVENTION: DNA SEQUENCES FOR STRAIN ANALYSIS IN MYCOBACTERIUM

; TITLE OF INVENTION: TUBERCULOSIS

; FILE REFERENCE: 24366-20007.00

; CURRENT APPLICATION NUMBER: US/09/103.840A

; CURRENT FILING DATE: 1998-06-24

; NUMBER OF SEQ ID NOS: 2

; SOFTWARE: PatentIn Ver. 2.1

; SEQ ID NO 1



```

; NUMBER OF SEQ ID NOS: 472
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 271
; LENGTH: 301
; TYPE: DNA
; ORGANISM: Homo sapien
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)...(301)
; OTHER INFORMATION: n = A,T,C or G
US-09-352-616A-271

Alignment Scores:
Pred. No.: 47.2 Length: 301
Score: 8.00 Matches: 8
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 1.09% Indels: 0
DB: 4 Gaps: 0

US-10-045-815-4 (1-736) x US-09-352-616A-271 (1-301)

QY 673 AlaLeuTrpPheThrLeuAspPro 680
Db 193 GCTCTGTGGTTCACCTTGGACCCA 170

RESULT 8
US-09-232-149A-271/c
; Sequence 271, Application US/09232149A
; Patent No. 6465611
; GENERAL INFORMATION:
; APPLICANT: Xu, Jiangchun
; APPLICANT: Dillon, Davin C.
; APPLICANT: Mitcham, Jennifer Lynn
; TITLE OF INVENTION: COMPOUNDS FOR IMMUNOTHERAPY OF PROSTATE
; FILE REFERENCE: 210121.427C6
; CURRENT APPLICATION NUMBER: US/09/232,149A
; CURRENT FILING DATE: 1999-01-15
; NUMBER OF SEQ ID NOS: 338
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 271
; LENGTH: 301
; TYPE: DNA
; ORGANISM: Homo sapien
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)...(301)
; OTHER INFORMATION: n = A,T,C or G
US-09-232-149A-271

Alignment Scores:
Pred. No.: 47.2 Length: 301
Score: 8.00 Matches: 8
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 1.09% Indels: 0
DB: 4 Gaps: 0

US-10-045-815-4 (1-736) x US-09-232-149A-271 (1-301)

QY 673 AlaLeuTrpPheThrLeuAspPro 680
Db 193 GCTCTGTGGTTCACCTTGGACCCA 170

RESULT 9
US-09-159-812-271/c
; Sequence 271, Application US/09159812A
; Patent No. 6613872
; GENERAL INFORMATION:
; APPLICANT: Xu, Jiangchun
; APPLICANT: Dillon, Davin C.
; TITLE OF INVENTION: COMPOUNDS FOR IMMUNOTHERAPY OF
; FILE REFERENCE: 210121.427C6
; CURRENT APPLICATION NUMBER: US/09/159,812A
; CURRENT FILING DATE: 1998-09-23
; NUMBER OF SEQ ID NOS: 306
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 271
; LENGTH: 301
; TYPE: DNA
; ORGANISM: Homo sapien
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)...(301)
; OTHER INFORMATION: n = A,T,C or G
US-09-159-812-271

Alignment Scores:
Pred. No.: 47.2 Length: 301
Score: 8.00 Matches: 8
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 1.09% Indels: 0
DB: 4 Gaps: 0

US-10-045-815-4 (1-736) x US-09-159-812-271 (1-301)

QY 673 AlaLeuTrpPheThrLeuAspPro 680
Db 193 GCTCTGTGGTTCACCTTGGACCCA 170

RESULT 10
US-09-636-215-271/c
; Sequence 271, Application US/09636215
; Patent No. 6620922
; GENERAL INFORMATION:
; APPLICANT: Xu, Jiangchun
; APPLICANT: Dillon, Davin C.
; APPLICANT: Mitcham, Jennifer L.
; APPLICANT: Harlocker, Susan L.
; APPLICANT: Jiang, Yuqui
; APPLICANT: Henderson, Robert A.
; APPLICANT: Kalos, Michael D.
; APPLICANT: Fanger, Gary R.
; APPLICANT: Retter, Marc W.
; APPLICANT: Stolk, John A.
; APPLICANT: Day, Craig H.
; APPLICANT: Vedvick, Thomas S.
; APPLICANT: Carter, Darrick
; APPLICANT: Li, Samuel
; APPLICANT: Wang, Aljun
; APPLICANT: Skeiky, Yasir A.W.
; APPLICANT: Hepler, William
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY
; FILE REFERENCE: 210121.42717C17
; CURRENT APPLICATION NUMBER: US/09/636,215
; CURRENT FILING DATE: 2000-08-10
; NUMBER OF SEQ ID NOS: 852
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 271
; LENGTH: 301
; TYPE: DNA
; ORGANISM: Homo sapien
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)...(301)
; OTHER INFORMATION: n = A,T,C or G
US-09-636-215-271

Alignment Scores:
Pred. No.: 47.2 Length: 301
Score: 8.00 Matches: 8
Percent Similarity: 100.00% Conservative: 0
; FILE REFERENCE: 210121.428C5
; CURRENT APPLICATION NUMBER: US/09/159,812A
; CURRENT FILING DATE: 1998-09-23
; NUMBER OF SEQ ID NOS: 306
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 271
; LENGTH: 301
; TYPE: DNA
; ORGANISM: Homo sapien
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)...(301)
; OTHER INFORMATION: n = A,T,C or G
US-09-159-812-271

Alignment Scores:
Pred. No.: 47.2 Length: 301
Score: 8.00 Matches: 8
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 1.09% Indels: 0
DB: 4 Gaps: 0

US-10-045-815-4 (1-736) x US-09-159-812-271 (1-301)

QY 673 AlaLeuTrpPheThrLeuAspPro 680
Db 193 GCTCTGTGGTTCACCTTGGACCCA 170

RESULT 11
US-09-636-215-271/c
; Sequence 271, Application US/09636215
; Patent No. 6620922
; GENERAL INFORMATION:
; APPLICANT: Xu, Jiangchun
; APPLICANT: Dillon, Davin C.
; APPLICANT: Mitcham, Jennifer L.
; APPLICANT: Harlocker, Susan L.
; APPLICANT: Jiang, Yuqui
; APPLICANT: Henderson, Robert A.
; APPLICANT: Kalos, Michael D.
; APPLICANT: Fanger, Gary R.
; APPLICANT: Retter, Marc W.
; APPLICANT: Stolk, John A.
; APPLICANT: Day, Craig H.
; APPLICANT: Vedvick, Thomas S.
; APPLICANT: Carter, Darrick
; APPLICANT: Li, Samuel
; APPLICANT: Wang, Aljun
; APPLICANT: Skeiky, Yasir A.W.
; APPLICANT: Hepler, William
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY
; FILE REFERENCE: 210121.42717C17
; CURRENT APPLICATION NUMBER: US/09/636,215
; CURRENT FILING DATE: 2000-08-10
; NUMBER OF SEQ ID NOS: 852
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 271
; LENGTH: 301
; TYPE: DNA
; ORGANISM: Homo sapien
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)...(301)
; OTHER INFORMATION: n = A,T,C or G
US-09-636-215-271

Alignment Scores:
Pred. No.: 47.2 Length: 301
Score: 8.00 Matches: 8
Percent Similarity: 100.00% Conservative: 0
; FILE REFERENCE: 210121.428C5
; CURRENT APPLICATION NUMBER: US/09/159,812A
; CURRENT FILING DATE: 1998-09-23
; NUMBER OF SEQ ID NOS: 306
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 271
; LENGTH: 301
; TYPE: DNA
; ORGANISM: Homo sapien
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)...(301)
; OTHER INFORMATION: n = A,T,C or G
US-09-159-812-271

Alignment Scores:
Pred. No.: 47.2 Length: 301
Score: 8.00 Matches: 8
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 1.09% Indels: 0
DB: 4 Gaps: 0

US-10-045-815-4 (1-736) x US-09-159-812-271 (1-301)

QY 673 AlaLeuTrpPheThrLeuAspPro 680
Db 193 GCTCTGTGGTTCACCTTGGACCCA 170

RESULT 12
US-09-636-215-271/c
; Sequence 271, Application US/09636215
; Patent No. 6620922
; GENERAL INFORMATION:
; APPLICANT: Xu, Jiangchun
; APPLICANT: Dillon, Davin C.
; APPLICANT: Mitcham, Jennifer L.
; APPLICANT: Harlocker, Susan L.
; APPLICANT: Jiang, Yuqui
; APPLICANT: Henderson, Robert A.
; APPLICANT: Kalos, Michael D.
; APPLICANT: Fanger, Gary R.
; APPLICANT: Retter, Marc W.
; APPLICANT: Stolk, John A.
; APPLICANT: Day, Craig H.
; APPLICANT: Vedvick, Thomas S.
; APPLICANT: Carter, Darrick
; APPLICANT: Li, Samuel
; APPLICANT: Wang, Aljun
; APPLICANT: Skeiky, Yasir A.W.
; APPLICANT: Hepler, William
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY
; FILE REFERENCE: 210121.42717C17
; CURRENT APPLICATION NUMBER: US/09/636,215
; CURRENT FILING DATE: 2000-08-10
; NUMBER OF SEQ ID NOS: 852
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 271
; LENGTH: 301
; TYPE: DNA
; ORGANISM: Homo sapien
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)...(301)
; OTHER INFORMATION: n = A,T,C or G
US-09-636-215-271

Alignment Scores:
Pred. No.: 47.2 Length: 301
Score: 8.00 Matches: 8
Percent Similarity: 100.00% Conservative: 0
; FILE REFERENCE: 210121.428C5
; CURRENT APPLICATION NUMBER: US/09/159,812A
; CURRENT FILING DATE: 1998-09-23
; NUMBER OF SEQ ID NOS: 306
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 271
; LENGTH: 301
; TYPE: DNA
; ORGANISM: Homo sapien
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)...(301)
; OTHER INFORMATION: n = A,T,C or G
US-09-159-812-271

Alignment Scores:
Pred. No.: 47.2 Length: 301
Score: 8.00 Matches: 8
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 1.09% Indels: 0
DB: 4 Gaps: 0

US-10-045-815-4 (1-736) x US-09-159-812-271 (1-301)

QY 673 AlaLeuTrpPheThrLeuAspPro 680
Db 193 GCTCTGTGGTTCACCTTGGACCCA 170

RESULT 13
US-09-636-215-271/c
; Sequence 271, Application US/09636215
; Patent No. 6620922
; GENERAL INFORMATION:
; APPLICANT: Xu, Jiangchun
; APPLICANT: Dillon, Davin C.
; APPLICANT: Mitcham, Jennifer L.
; APPLICANT: Harlocker, Susan L.
; APPLICANT: Jiang, Yuqui
; APPLICANT: Henderson, Robert A.
; APPLICANT: Kalos, Michael D.
; APPLICANT: Fanger, Gary R.
; APPLICANT: Retter, Marc W.
; APPLICANT: Stolk, John A.
; APPLICANT: Day, Craig H.
; APPLICANT: Vedvick, Thomas S.
; APPLICANT: Carter, Darrick
; APPLICANT: Li, Samuel
; APPLICANT: Wang, Aljun
; APPLICANT: Skeiky, Yasir A.W.
; APPLICANT: Hepler, William
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY
; FILE REFERENCE: 210121.42717C17
; CURRENT APPLICATION NUMBER: US/09/636,215
; CURRENT FILING DATE: 2000-08-10
; NUMBER OF SEQ ID NOS: 852
; SOFTWARE: FastSeq for Windows Version 3.0
;
```

Best Local Similarity: 100.00% Mismatches: 0  
Query Match: 1.09% Indels: 0  
DB: 4 Gaps: 0

US-10-045-815-4 (1-736) x US-09-636-215-271 (1-301)

Qy 673 AlaLeuTrpPheThrLeuAspPro 680  
Db 193 GCTCTGTGGTTCACCTTGGACCCA 170

RESULT 11

US-09-685-166A-271/c  
; Sequence 271, Application US/09685166A  
; Patent No. 6630305  
; GENERAL INFORMATION:  
; APPLICANT: Xu, Jiangchun  
; APPLICANT: Dillon, Davin C.  
; APPLICANT: Mitcham, Jennifer L.  
; APPLICANT: Harlocker, Susan L.  
; APPLICANT: Jiang, Yuqi  
; APPLICANT: Henderson, Robert A.  
; APPLICANT: Kalos, Michael D.  
; APPLICANT: Fanger, Gary R.  
; APPLICANT: Retter, Marc W.  
; APPLICANT: Stolk, John A.  
; APPLICANT: Day, Craig H.  
; APPLICANT: Vedvick, Thomas S.  
; APPLICANT: Carter, Darrick  
; APPLICANT: Li, Samuel  
; APPLICANT: Wang, Aijun  
; APPLICANT: Skeiky, Yasir A.W.  
; APPLICANT: Hepler, William  
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY AND  
; TITLE OF INVENTION: DIAGNOSIS OF PROSTATE CANCER  
; FILE REFERENCE: 210121.427C21  
; CURRENT APPLICATION NUMBER: US/09/685.166A  
; CURRENT FILING DATE: 2000-10-10  
; NUMBER OF SEQ ID NOS: 898  
; SOFTWARE: FastSeq for Windows Version 3.0  
; SEQ ID NO 271  
; LENGTH: 301  
; TYPE: DNA  
; ORGANISM: Homo sapien  
; FEATURE:  
; NAME/KEY: misc feature  
; LOCATION: (1)...(301)  
; OTHER INFORMATION: n = A,T,C or G

US-09-685-166A-271

Alignment Scores:  
Pred. No.: 47.2 Length: 301  
Score: 8.00 Matches: 8  
Percent Similarity: 100.00% Conservative: 0  
Best Local Similarity: 100.00% Mismatches: 0  
Query Match: 1.09% Indels: 0  
DB: 4 Gaps: 0

US-10-045-815-4 (1-736) x US-09-685-166A-271 (1-301)

Qy 673 AlaLeuTrpPheThrLeuAspPro 680  
Db 193 GCTCTGTGGTTCACCTTGGACCCA 170

RESULT 12

US-09-688-489-271/c  
; Sequence 271, Application US/09688489  
; Patent No. 6664377  
; GENERAL INFORMATION:  
; APPLICANT: Xu, Jiangchun  
; APPLICANT: Dillon, Davin C.  
; APPLICANT: Mitcham, Jennifer Lynn  
; TITLE OF INVENTION: COMPOUNDS FOR IMMUNOTHERAPY OF PROSTATE  
; TITLE OF INVENTION: CANCER AND METHODS FOR THEIR USE

; FILE REFERENCE: 210121.427D2  
; CURRENT APPLICATION NUMBER: US/09/688.489  
; CURRENT FILING DATE: 2000-10-13  
; NUMBER OF SEQ ID NOS: 338  
; SOFTWARE: FastSeq for Windows Version 3.0  
; SEQ ID NO 271  
; LENGTH: 301  
; TYPE: DNA  
; ORGANISM: Homo sapien  
; FEATURE:  
; NAME/KEY: misc feature  
; LOCATION: (1)...(301)  
; OTHER INFORMATION: n = A,T,C or G

US-09-688-489-271

Alignment Scores:  
Pred. No.: 47.2 Length: 301  
Score: 8.00 Matches: 8  
Percent Similarity: 100.00% Conservative: 0  
Best Local Similarity: 100.00% Mismatches: 0  
Query Match: 1.09% Indels: 0  
DB: 4 Gaps: 0

US-10-045-815-4 (1-736) x US-09-688-489-271 (1-301)

Qy 673 AlaLeuTrpPheThrLeuAspPro 680  
Db 193 GCTCTGTGGTTCACCTTGGACCCA 170

RESULT 13

US-08-407-620A-12/c  
; Sequence 12, Application US/08407620A  
; Patent No. 6569430  
; GENERAL INFORMATION:  
; APPLICANT: WALDMANN, HERMAN  
; APPLICANT: CLARK, MICHAEL R.  
; APPLICANT: WINTER, GREGORY P.  
; APPLICANT: RIECEMANN, LUTZ  
; TITLE OF INVENTION: ANTIBODIES  
; NUMBER OF SEQUENCES: 53  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: NIXON & VANDERHUYE P.C.  
; STREET: 1100 NORTH GLEBE ROAD, 8TH FLOOR  
; CITY: ARLINGTON  
; STATE: VIRGINIA  
; COUNTRY: U.S.A.  
; ZIP: 22201-4714  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/407,620A  
; FILING DATE: 21-MAR-1995  
; CLASSIFICATION: 424  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 08/235,705  
; FILING DATE: 29-APR-1994  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 08/099,480  
; FILING DATE: 30-JUL-1993  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 07/921,601  
; FILING DATE: 03-AUG-1992  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 07/424,233  
; FILING DATE: 12-OCT-1989  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: GB 88036228  
; FILING DATE: 12-FEB-1988  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: GB 8804464

```
/ FILING DATE: 25-FEB-1988
/ ATTORNEY/AGENT INFORMATION:
/ NAME: MITCHARD, LEONARD C.
/ REGISTRATION NUMBER: 29,009
/ REFERENCE/DOCKET NUMBER: 604-325
/ TELECOMMUNICATION INFORMATION:
/ TELEPHONE: (703) 816-4000
/ TELEFAX: (703) 816-4100
/ TELEX: 200797 NIXN UR
/ INFORMATION FOR SEQ ID NO: 12:
/ SEQUENCE CHARACTERISTICS:
/ LENGTH: 363 base pairs
/ TYPE: nucleic acid
/ STRANDEDNESS: single
/ TOPOLOGY: linear
/ MOLECULE TYPE: DNA (genomic)
US-08-407-620A-12

Alignment Scores:
Pred. No.: 56.5 Length: 363
Score: 8.00 Matches: 8
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 1.09% Indels: 0
DB: 4 Gaps: 0

US-10-045-815-4 (1-736) x US-08-407-620A-12 (1-363)
QY 718 ProGluProAlaGlnGluSerLeu 725
Db 77 CCAGAACCTGCACAGAGAGCTC 54

RESULT 14
US-09-252-991A-4191
/ Sequence 4191, Application US/09252991A
/ Patent No. 6551795
/ GENERAL INFORMATION:
/ APPLICANT: Marc J. Rubenfield et al.
/ TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
/ FILE REFERENCE: 107196.136
/ CURRENT APPLICATION NUMBER: US/09/252,991A
/ PRIOR FILING DATE: 1999-02-18
/ PRIOR APPLICATION NUMBER: US 60/074,788
/ PRIOR FILING DATE: 1998-02-18
/ PRIOR APPLICATION NUMBER: US 60/094,190
/ PRIOR FILING DATE: 1998-07-27
/ NUMBER OF SEQ ID NOS: 33142
/ SEQ ID NO 4191
/ LENGTH: 375
/ TYPE: DNA
/ ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-4191

Alignment Scores:
Pred. No.: 58.3 Length: 375
Score: 8.00 Matches: 8
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 1.09% Indels: 0
DB: 4 Gaps: 0

US-10-045-815-4 (1-736) x US-09-252-991A-4191 (1-375)
QY 60 ArgAlaLeuArgSerArgAlaAla 67
Db 212 CGGCGGCTTCGATCAGCGCGCG 235

RESULT 15
US-09-582-337-21/C
/ Sequence 21, Application US/09582337
/ Patent No. 6562618
/ GENERAL INFORMATION:
/ APPLICANT: Japan Tobacco, Inc.
/ TITLE OF INVENTION: Monoclonal Antibody Against Connective Tissue Growth Factor
/ FILE REFERENCE: J1-009PCT
/ CURRENT APPLICATION NUMBER: US/09/582,337
/ CURRENT FILING DATE: 2000-06-23
/ PRIOR APPLICATION NUMBER: JP P1997-367699
/ PRIOR FILING DATE: 1997-12-25
/ PRIOR APPLICATION NUMBER: JP P1998-356183
/ PRIOR FILING DATE: 1998-12-15
/ NUMBER OF SEQ ID NOS: 27
/ SOFTWARE: PatentIn Ver. 2.0
/ SEQ ID NO 21
/ LENGTH: 387
/ TYPE: DNA
/ ORGANISM: Homo sapiens
/ FEATURE:
/ NAME/KEY: CDS
/ LOCATION: (1)..(387)
/ NAME/KEY: sig_peptide
/ LOCATION: (1)..(47)
/ OTHER INFORMATION: Initiation codon and a portion of a signal
/ OTHER INFORMATION: sequence are lacked.
/ NAME/KEY: V region
/ LOCATION: (48)..(335)
US-09-582-337-21

Alignment Scores:
Pred. No.: 60.1 Length: 387
Score: 8.00 Matches: 8
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 1.09% Indels: 0
DB: 4 Gaps: 0

US-10-045-815-4 (1-736) x US-09-582-337-21 (1-387)
QY 651 PheSerSerGlyThrGluAsnPro 658
Db 50 TTTCATCCGCGACCGAACCACCA 27

Search completed: July 19, 2004, 00:22:40
Job time : 3383 secs
```

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: July 18, 2004, 09:43:04 ; Search time 17 Seconds  
(without alignments)  
2254.331 Million cell updates/sec

Title: US-10-045-815-4  
Perfect score: 736  
Sequence: 1 MAVRALKLLTLLAVVAAS.....PPEPAQESLSGSESKPKDEL 736

Scoring table: OLIGO  
Gapop 60.0 , Gapext 60.0

Searched: 141681 seqs, 52070155 residues

Word size : 0

Total number of hits satisfying chosen parameters: 141681

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database : SwissProt\_42:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	9	1.2	75	1 TRBK_AGRU	P54912 agrobacteri
2	9	1.2	364	1 DRG2_HUMAN	P55039 homo sapien
3	9	1.2	364	1 DRG2_MOUSE	O9qxb9 mus musculu
4	8	1.1	162	1 CAV2_FUGRU	O9ygm9 fugu rubrip
5	8	1.1	401	1 CASP_HUMAN	O75718 homo sapien
6	8	1.1	431	1 SC65_RAT	O64375 rattus norv
7	8	1.1	437	1 NO55_HUMAN	O92791 homo sapien
8	8	1.1	493	1 UDPE_NPVCD	O90358 chorisoneu
9	8	1.1	508	1 OTOL_ONCKE	P83371 oncorhynch
10	7	1.0	84	1 IM13_ORYSA	O9xgv5 oryza sativ
11	7	1.0	103	1 YKN7_YEAST	P36064 saccharomyc
12	7	1.0	121	1 TGFA_MACMU	P55244 macaca mula
13	7	1.0	126	1 RL7_CHLTE	O8k216 chlorobium
14	7	1.0	133	1 TGFA_SHEEP	P98135 ovis aries
15	7	1.0	139	1 TGFA_MOUSE	P48030 mus musculu
16	7	1.0	159	1 TGFA_RAT	P01134 rattus norv
17	7	1.0	160	1 TGFA_HUMAN	P01135 homo sapien
18	7	1.0	160	1 TGFA_PIG	O06922 sus scrofa
19	7	1.0	183	1 HSLV_XANCP	O8p551 xanthomonas
20	7	1.0	196	1 RR4_ALOPR	P36445 alopecurus
21	7	1.0	196	1 RR4_ANDIS	P36446 andropogon
22	7	1.0	196	1 RR4_ANTOD	P36447 anthoxanthu
23	7	1.0	196	1 RR4_ARUCA	P36448 arundinaria
24	7	1.0	196	1 RR4_BRAPI	P36449 brachypodiu
25	7	1.0	196	1 RR4_BROER	P36451 bromus erer
26	7	1.0	196	1 RR4_CALEP	P36452 calamagrost
27	7	1.0	196	1 RR4_DENGI	P36454 dendrocalam
28	7	1.0	196	1 RR4_ELEIN	P36456 eleusine in
29	7	1.0	196	1 RR4_ELICA	P36457 elymus cana
30	7	1.0	196	1 RR4_FESGI	P36458 festuca gig
31	7	1.0	196	1 RR4_FESPR	P36459 festuca pra
32	7	1.0	196	1 RR4_LYGSP	P36464 lygeum spar
33	7	1.0	196	1 RR4_MELAL	P36465 melica alti

RESULT 1				
ID	TRBK_AGRU	STANDARD;	PRT;	75 AA.
AC	P54912;			
DT	01-OCT-1996 (Rel. 34, Created)			
DT	01-OCT-1996 (Rel. 34, Last sequence update)			
DT	16-OCT-2001 (Rel. 40, Last annotation update)			
DE	Conjugal transfer protein trbK precursor.			
GN	TRBK.			
OS	Agrobacterium tumefaciens.			
OG	Plasmid pTiA6NC.			
OC	Bacteria; Proteobacteria; Alphaproteobacteria; Rhizobiales;			
OC	Rhizobiaceae; Rhizobium/Agrobacterium group; Agrobacterium.			
OX	NCBI_TaxID=358;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RX	MEDLINE=96312368; PubMed=8763954;			
RA	Alt-Morbe J., Stryker J.L., Fuqua C., Li P.L., Farrand S.K.,			
RA	Winans S.C.;			
RT	"The conjugal transfer system of Agrobacterium tumefaciens			
RT	octopine-type Ti plasmids is closely related to the transfer system			
RT	of an IncP plasmid and distantly related to Ti plasmid vir genes.";			
RL	J. Bacteriol. 178:4248-4257(1996).			
CC	-----			
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration			
CC	between the Swiss Institute of Bioinformatics and the EMBL outstation -			
CC	the European Bioinformatics Institute. There are no restrictions on its			
CC	use by non-profit institutions as long as its content is in no way			
CC	modified and this statement is not removed. Usage by and for commercial			
CC	entities requires a license agreement (See <a href="http://www.isb-sib.ch/announce/">http://www.isb-sib.ch/announce/</a>			
CC	or send an email to <a href="mailto:license@isb-sib.ch">license@isb-sib.ch</a> ).			
CC	-----			
DR	EMBL; AF242881; AAB95099.1; -			
KW	Conjugation; Plasmid; Signal.			
FT	SIGNAL	1	20	POTENTIAL.
FT	CHAIN	21	75	CONJUGAL TRANSFER PROTEIN TRBK.
SQ	SEQUENCE	75 AA;	8023 MW;	E00C5114C0497027 CRC64;
Query Match				
Best Local Similarity 1.2%; Score 9; DB 1; Length 75;				
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;				
QY	12 LLAVVVAAS 20			
DB	11 LLAVVVAAS 19			
-----				
RESULT 2				
DRG2_HUMAN				
ID	DRG2_HUMAN	STANDARD;	PRT;	364 AA.
AC	P55039; Q9BW82;			
DT	01-OCT-1996 (Rel. 34, Created)			
DT	01-OCT-1996 (Rel. 34, Last sequence update)			
DT	10-OCT-2003 (Rel. 42, Last annotation update)			
DE	Developmentally regulated GTP-binding protein 2 (DRG 2).			
GN	DRG2.			

P36466 melica unif  
P36467 pennisetum  
P36471 setaria vir  
P36472 stipa capen  
P36473 tragus race  
P47149 saccharomyc  
P02355 zea mays (m  
P12147 oryza sativ  
Q95h61 triticum ae  
Q9tly7 cyanidium c  
Q8ug34 agrobacteri  
Q8m9w2 chaetosphae

ALIGNMENTS

OS Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
NCBI\_TaxID=9606;  
OX  
RN  
RP  
RQ  
RX  
RY  
RZ  
SA  
SB  
SC  
SD  
SE  
SF  
SG  
SH  
SI  
SJ  
SK  
SL  
SM  
SN  
SO  
SQ  
ST  
SU  
SV  
SW  
SX  
SY  
SZ  
TA  
TB  
TC  
TD  
TE  
TF  
TG  
TH  
TI  
TJ  
TK  
TL  
TM  
TN  
TO  
TP  
TP  
TR  
TS  
TT  
TU  
TV  
TW  
TX  
TY  
TZ  
UA  
UB  
UC  
UD  
UE  
UF  
UG  
UH  
UI  
UJ  
UK  
UL  
UM  
UN  
UO  
UP  
UQ  
UR  
US  
UT  
UU  
UV  
UW  
UX  
UY  
UZ  
VA  
VB  
VC  
VD  
VE  
VF  
VG  
VH  
VI  
VJ  
VK  
VL  
VM  
VN  
VO  
VP  
VQ  
VR  
VS  
VT  
VU  
VV  
VV  
VW  
VX  
VY  
VZ  
WA  
WB  
WC  
WD  
WE  
WF  
WG  
WH  
WI  
WJ  
WK  
WL  
WM  
WN  
WO  
WP  
WQ  
WR  
WS  
WT  
WU  
WV  
WW  
WX  
WY  
WZ  
XA  
XB  
XC  
XD  
XE  
XF  
XG  
XH  
XI  
XJ  
XK  
XL  
XM  
XN  
XO  
XP  
XQ  
XR  
XS  
XT  
XU  
XV  
XW  
XX  
XY  
XZ  
YA  
YB  
YC  
YD  
YE  
YF  
YG  
YH  
YI  
YJ  
YK  
YL  
YM  
YN  
YO  
YP  
YQ  
YR  
YS  
YT  
YU  
YV  
YW  
YX  
YY  
YZ  
ZA  
ZB  
ZC  
ZD  
ZE  
ZF  
ZG  
ZH  
ZI  
ZJ  
ZK  
ZL  
ZM  
ZN  
ZO  
ZP  
ZQ  
ZR  
ZS  
ZT  
ZU  
ZV  
ZW  
ZX  
ZY  
ZZ

Query Match 1.2%; Score 9; DB 1; Length 364;  
Best Local Similarity 100.0%; Pred. No. 0.78;  
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 367 QRSLLLEKEL 375  
Db 155 QRSLLLEKEL 163

RESULT 3  
DRG2\_MOUSE  
ID DRG2\_MOUSE STANDARD; PRT; 364 AA.  
AC Q9QXB9;  
DT 16-OCT-2001 (Rel. 40, Created)  
DT 16-OCT-2001 (Rel. 40, Last sequence update)  
DT 16-OCT-2001 (Rel. 40, Last annotation update)  
DE Developmentally regulated GTP-binding protein 2 (DRG 2).  
DN  
GN Mus musculus (Mouse).  
OS Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
OX NCBI\_TaxID=10090;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=202325460; PubMed=10760581;  
RA Li B., Trueb B.;  
RT "DRG represents a family of two closely related GTP-binding  
proteins";  
RL Biochim. Biophys. Acta 1491:196-204(2000).  
CC -!- FUNCTION: May play a role in cell proliferation, differentiation  
and death.  
CC -!- SUBCELLULAR LOCATION: Cytoplasmic (By similarity).  
CC -!- TISSUE SPECIFICITY: FAIRLY HIGH LEVELS IN LIVER, HEART, KIDNEY,  
AND BRAIN. VERY LOW LEVELS IN LUNG, SPLEEN, TESTIS AND SKELETAL  
MUSCLE.  
CC -!- SIMILARITY: Belongs to the GTP1 / OBG family.  
CC  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
between the Swiss Institute of Bioinformatics and the EMBL outstation -  
the European Bioinformatics Institute. There are no restrictions on its  
use by non-profit institutions as long as its content is in no way  
modified and this statement is not removed. Usage by and for commercial  
entities requires a license agreement (See <http://www.isb-sib.ch/announcement/>  
or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC EMBL; AJ243590; CAB65258.1; -  
DR MGB; MG1:1342307; Drg2.  
DR InterPro; IPR006074; GTP1/OBG dom.  
DR InterPro; IPR006073; GTP1\_OBG.  
DR InterPro; IPR006169; GTP1\_OBG sub.  
DR InterPro; IPR005225; Small GTP.  
DR InterPro; IPR004095; TGS dom.  
DR Pfam; PF01018; GTP1\_OBG; 1.  
DR Pfam; PF02824; TGS; 1.  
DR PRINTS; PR00326; GTP1\_OBG.  
DR TIGRFAMs; TIGR00231; small\_gtp; 1.  
DR PROSITE; PS00905; GTP1\_OBG; 1.  
KW GTP-binding. 69 76 GTP (BY SIMILARITY).  
FT NP\_BIND 115 119 GTP (BY SIMILARITY).  
FT NP\_BIND 246 249 GTP (BY SIMILARITY).  
SQ SEQUENCE 364 AA; 40718 MW; 932353C8FF257BE8 CRC64;  
Query Match 1.2%; Score 9; DB 1; Length 364;  
Best Local Similarity 100.0%; Pred. No. 0.78;  
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 367 QRSLLLEKEL 375  
Db 155 QRSLLLEKEL 163

RESULT 3  
DRG2\_MOUSE  
ID DRG2\_MOUSE STANDARD; PRT; 364 AA.  
AC Q9QXB9;  
DT 16-OCT-2001 (Rel. 40, Created)  
DT 16-OCT-2001 (Rel. 40, Last sequence update)  
DT 16-OCT-2001 (Rel. 40, Last annotation update)  
DE Developmentally regulated GTP-binding protein 2 (DRG 2).  
DN  
GN Mus musculus (Mouse).  
OS Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
OX NCBI\_TaxID=10090;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=202325460; PubMed=10760581;  
RA Li B., Trueb B.;  
RT "DRG represents a family of two closely related GTP-binding  
proteins";  
RL Biochim. Biophys. Acta 1491:196-204(2000).  
CC -!- FUNCTION: May play a role in cell proliferation, differentiation  
and death.  
CC -!- SUBCELLULAR LOCATION: Cytoplasmic (By similarity).  
CC -!- TISSUE SPECIFICITY: FAIRLY HIGH LEVELS IN LIVER, HEART, KIDNEY,  
AND BRAIN. VERY LOW LEVELS IN LUNG, SPLEEN, TESTIS AND SKELETAL  
MUSCLE.  
CC -!- SIMILARITY: Belongs to the GTP1 / OBG family.  
CC  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
between the Swiss Institute of Bioinformatics and the EMBL outstation -  
the European Bioinformatics Institute. There are no restrictions on its  
use by non-profit institutions as long as its content is in no way  
modified and this statement is not removed. Usage by and for commercial  
entities requires a license agreement (See <http://www.isb-sib.ch/announcement/>  
or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC EMBL; X80754; CAH56730.1; -  
DR EMBL; BC000493; AAH00493.1; -  
DR FIR; A55014; A55014.  
DR Genew; HGNC:3030; DRG2.  
DR MIM; 602986; -  
DR GO; GO:0005525; P:GTP binding; TAS.  
DR GO; GO:0008151; P:cell growth and/or maintenance; TAS.  
DR GO; GO:0007165; P:signal transduction; TAS.  
DR InterPro; IPR006074; GTP1/OBG dom.  
DR InterPro; IPR006073; GTP1\_OBG.  
DR InterPro; IPR006169; GTP1\_OBG sub.  
DR InterPro; IPR005225; Small GTP.  
DR InterPro; IPR004095; TGS dom.  
DR Pfam; PF01018; GTP1\_OBG; 1.  
DR Pfam; PF02824; TGS; 1.  
DR PRINTS; PR00326; GTP1\_OBG.  
DR TIGRFAMs; TIGR00231; small\_gtp; 1.  
DR PROSITE; PS00905; GTP1\_OBG; 1.  
KW GTP-binding. 69 76 GTP (BY SIMILARITY).  
FT NP\_BIND 115 119 GTP (BY SIMILARITY).  
FT NP\_BIND 246 249 GTP (BY SIMILARITY).  
SQ SEQUENCE 364 AA; 40718 MW; 932353C8FF257BE8 CRC64;  
Query Match 1.2%; Score 9; DB 1; Length 364;  
Best Local Similarity 100.0%; Pred. No. 0.78;  
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 367 QRSLLLEKEL 375  
Db 155 QRSLLLEKEL 163

RESULT 3  
DRG2\_MOUSE  
ID DRG2\_MOUSE STANDARD; PRT; 364 AA.  
AC Q9QXB9;  
DT 16-OCT-2001 (Rel. 40, Created)  
DT 16-OCT-2001 (Rel. 40, Last sequence update)  
DT 16-OCT-2001 (Rel. 40, Last annotation update)  
DE Developmentally regulated GTP-binding protein 2 (DRG 2).  
DN  
GN Mus musculus (Mouse).  
OS Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
OX NCBI\_TaxID=10090;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=202325460; PubMed=10760581;  
RA Li B., Trueb B.;  
RT "DRG represents a family of two closely related GTP-binding  
proteins";  
RL Biochim. Biophys. Acta 1491:196-204(2000).  
CC -!- FUNCTION: May play a role in cell proliferation, differentiation  
and death.  
CC -!- SUBCELLULAR LOCATION: Cytoplasmic (By similarity).  
CC -!- TISSUE SPECIFICITY: FAIRLY HIGH LEVELS IN LIVER, HEART, KIDNEY,  
AND BRAIN. VERY LOW LEVELS IN LUNG, SPLEEN, TESTIS AND SKELETAL  
MUSCLE.  
CC -!- SIMILARITY: Belongs to the GTP1 / OBG family.  
CC  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
between the Swiss Institute of Bioinformatics and the EMBL outstation -  
the European Bioinformatics Institute. There are no restrictions on its  
use by non-profit institutions as long as its content is in no way  
modified and this statement is not removed. Usage by and for commercial  
entities requires a license agreement (See <http://www.isb-sib.ch/announcement/>  
or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC EMBL; X80754; CAH56730.1; -  
DR EMBL; BC000493; AAH00493.1; -  
DR FIR; A55014; A55014.  
DR Genew; HGNC:3030; DRG2.  
DR MIM; 602986; -  
DR GO; GO:0005525; P:GTP binding; TAS.  
DR GO; GO:0008151; P:cell growth and/or maintenance; TAS.  
DR GO; GO:0007165; P:signal transduction; TAS.  
DR InterPro; IPR006074; GTP1/OBG dom.  
DR InterPro; IPR006073; GTP1\_OBG.  
DR InterPro; IPR006169; GTP1\_OBG sub.  
DR InterPro; IPR005225; Small GTP.  
DR InterPro; IPR004095; TGS dom.  
DR Pfam; PF01018; GTP1\_OBG; 1.  
DR Pfam; PF02824; TGS;



Db 155 QRSLEKEL 163

RESULT 4

CAV2\_FUGRU STANDARD; PRT; 162 AA.

AC Q9YGM9;

DT 16-OCT-2001 (Rel. 40, Created)

DT 16-OCT-2001 (Rel. 40, Last sequence update)

DT 15-MAR-2004 (Rel. 43, Last annotation update)

DE Caveolin-2.

GN CAV2 OR CAV-2.

OS Fugu rubripes (Japanese pufferfish) (Takifugu rubripes).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

OC Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neteleostei;

OC Acanthomorpha; Acanthopterygii; Percormorpha; Tetraodontiformes;

OC Tetraodontidae; Tetraodontidae; Takifugu.

OX NCBI\_TaxID=31033;

[1]

RN SEQUENCE FROM N.A.

RP Cottage A.J.;

RA Submitted (AUG-1998) to the EMBL/GenBank/DBJ databases.

RL [2]

RX MEDLINE=22799194; PubMed=12917688;

RA Thomas J.W., Touchman J.W., Blakesley R.W., Bouffard G.G., Beckstrom-Sternberg S.M., Margulies E.H., Blanchette M., Siepel A.C., Thomas P.J., McDowell J.C., Maskeri B., Hansen N.F., Schwartz M.S., Weber R.J., Kent W.J., Karolchik D., Bruen T.C., Bevan R., Cutler D.J., Schwartz S., Elmitaki L., Idol J.R., Prasad A.B., Lee-Lin S.-Q., Maduro V.V., Summers T.J., Portnoy M.E., Dietrich N.L., Akhtar N., Ayele K., Benjamin B., Carliaga K., Brinkley C.P., Brooks S.Y., Granits E., Guan X., Gupta J., Haghighi P., Ho S.-L., Huang M.C., Karlins E., Laric P.L., Legaspi R., Lim M.J., Maduro Q.L., Masinello C.A., Mastrian S.D., McCloskey J.C., Pearson R., Stantripo S., Tingson E.E., Tran J.T., Tsugeon C., Vogt J.L., Walker M.A., Wetherby K.D., Wiggins L.S., Young A.C., Zhang L.-H., Osoegawa K., Zhu B., Zhao B., Shu C.L., De Jong P.J., Lawrence C.E., Smit A.F., Chakravarti A., Haussler D., Green P., Miller W., Green E.D.;

RA "Comparative analyses of multi-species sequences from targeted genomic regions.";

RT Nature 424:788-793(2003).

RL

CC -!- FUNCTION: May act as a scaffolding protein within caveolar membranes. Interacts directly with G-protein alpha subunits and can functionally regulate their activity (By similarity).

CC

CC -!- SUBUNIT: Homooligomer (By similarity).

CC

CC -!- SUBCELLULAR LOCATION: Membrane protein of caveolae. Potential hairpin-like structure in the membrane (By similarity).

CC

CC -!- SIMILARITY: Belongs to the caveolin family.

CC

CC

CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

CC

CC -----

CC EMBL; AJ010316; CAA09081.1; -

CC EMBL; AC090119; AAL40363.1; -

DR InterPro; IPR001612; Caveolin.

DR Pfam; PF01146; Caveolin; 1.

DR PROSITE; PS01210; CAVEOLIN; 1.

KW Transmembrane; Lipoprotein.

FT DOMAIN 1 86 CYTOPLASMIC (POTENTIAL).

FT TSANMEM 87 107 POTENTIAL.

FT DOMAIN 108 162 CYTOPLASMIC (POTENTIAL).

SQ SEQUENCE 162 AA; 18236 MW; 1D7CF4907D491253 CRC64;

Query Match 1.1%; Score 8; DB 1; Length 162;

Best Local Similarity 100.0%; Pred. No. 4.1;

Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 8 LTTLLAV 15

Db 87 LTTLLAV 94

RESULT 5

CASP\_HUMAN STANDARD; PRT; 401 AA.

ID \_CASP\_HUMAN

AC O757I8;

DT 16-OCT-2001 (Rel. 40, Created)

DT 16-OCT-2001 (Rel. 40, Last sequence update)

DT 10-OCT-2003 (Rel. 42, Last annotation update)

DE Cartilage-associated protein precursor.

GN CRTAP OR CASP

OS Homo sapiens (Human).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

OX NCBI\_TaxID=9606;

RN [1]

RN SEQUENCE FROM N.A.

RP TISSUE=Petal brain;

RC MEDLINE=20169181; PubMed=10703664;

RX Tonachini L., Morello R., Monticone M., Skaug J., Scherer S.W., Cancedda R., Castagnola P.;

RA "cDNA cloning, characterization and chromosome mapping of the gene encoding human cartilage associated protein (CRTAP).";

RT Cytogenet. Cell Genet. 87:191-194(1999).

RL [2]

RN SEQUENCE FROM N.A.

RP TISSUE=Kidney;

RC MEDLINE=22388257; PubMed=12477932;

RX Strausberg R.L., Feingold B.A., Grouse L.H., Derge J.G., Klausner R.D., Collins P.S., Wagner L., Shenmen C.M., Schuler G.D., Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K., Hopkins R.F., Jordan K., Moore T., Max S.I., Wang J., Hsieh F., Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L., Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E., Brownstein M.J., Ustin T.B., Toshiyuki S., Carninci P., Prange C., Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullahy S.J., Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H., Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W., Villalón D.K., Wuzny D.M., Sodergren E.J., Lu X., Gibbs R.A., Fahy J., Helton E., Kettman M., Madan A., Rodrigues S., Sanchez A., Whiting R.M., Touchman J.W., Green E.D., Dickinson G.G., Blakesley R.W., Grimwood J., Schmutz J., Myers R.M., Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M., Butterfield Y.S.N., Krzywinski M.I., Skalska U., Smalhus D.E., Schnerch A., Schein J.E., Jones S.J.M., Marra M.A.;

RA "Generation and initial analysis of more than 15,000 full-length human and mouse cDNA sequences.";

RT Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).

RL

CC -!- SUBCELLULAR LOCATION: Secreted. Extracellular matrix (By similarity).

CC

CC -!- TISSUE SPECIFICITY: Found in articular chondrocytes. Expressed in a variety of tissues.

CC

CC -!- SIMILARITY: BELONGS TO THE CRTAP / NO55 FAMILY.

CC

CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

CC

CC -----

CC EMBL; AJ006470; CAA07054.1; -

DR EMBL; BC008745; AAL08745.1; -

DR Genew; HGNC:2379; CRTAP.

DR MIM; 605497; -

DR InterPro; IPR008940; Prenyl trans.

DR Extracellular matrix; Signal.

KW SIGNAL 1 POTENTIAL.

FT CHAIN 27 401 CARTILAGE-ASSOCIATED PROTEIN.  
 FT CARBOHYD 87 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 363 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CONFLICT 137 E -> D (IN REF. 2).  
 SQ SEQUENCE 401 AA; 46561 MW; 4BEED4089195456F CRC64;

Query Match 1.1%; Score 8; DB 1; Length 401;  
 Best Local Similarity 100.0%; Pred. No. 9; Mismatches 0; Gaps 0;  
 Matches 8; Conservative 0; Indels 0;

Qy 110 FGGLLRRA 117  
 |||||  
 Db 109 FGGLLRRA 116

RESULT 6  
 SC65 RAT STANDARD; PRT; 431 AA.  
 ID Q64375;  
 DT 15-JUL-1998 (Rel. 36, Created)  
 DT 15-JUL-1998 (Rel. 36, Last sequence update)  
 DT 16-OCT-2001 (Rel. 40, Last annotation update)  
 DE Synaptonemal complex protein SC65.  
 GN SC65.  
 OS Rattus norvegicus (Rat).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.  
 OC NCBI\_TaxID=10116;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=Testis;  
 RX MEDLINE=93213429; PubMed=1363622;  
 RA Chen Q., Pearlman R.E., Moens P.B.;  
 RT "Isolation and characterization of a cDNA encoding a synaptonemal  
 complex protein.";  
 RL Biochem. Cell Biol. 70:1030-1038(1992).  
 CC -!- SUBCELLULAR LOCATION: Nuclear. Located in the pairing zone of the  
 synaptonemal complex.  
 CC -!- TISSUE SPECIFICITY: Found in testis, brain, heart and at a much  
 lower level in liver.  
 CC -!- SIMILARITY: BELONGS TO THE CRTAP / NO55 FAMILY.  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 the European Bioinformatics Institute. There are no restrictions on its  
 use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 CC EMBL; X65454; CAA6449.1; -;  
 CC F018; A56822; A56822.  
 CC InterPro; IPR008940; Prenyl\_trans.  
 KW Nuclear protein.  
 FT DOMAIN 319 328 GLU-RICH (ACIDIC).  
 FT DOMAIN 353 384 GLU-RICH (ACIDIC).  
 SQ SEQUENCE 431 AA; 49995 MW; 4A34F3029407B2E7 CRC64;

Query Match 1.1%; Score 8; DB 1; Length 431;  
 Best Local Similarity 100.0%; Pred. No. 9.6;  
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 116 RAACLRRC 123  
 |||||  
 Db 85 RAACLRRC 92

RESULT 7  
 NO55 HUMAN STANDARD; PRT; 437 AA.  
 AC Q92791; Q9H4F6;  
 DT 15-JUL-1998 (Rel. 36, Created)  
 DT 15-JUL-1998 (Rel. 36, Last sequence update)

DT 15-MAR-2004 (Rel. 43, Last annotation update)  
 DE Nuclear autoantigen No55.  
 GN SC65 OR NO55.  
 OC Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 OC NCBI\_TaxID=9606;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=Bladder epithelium;  
 RX MEDLINE=97015880; PubMed=8862517;  
 RA Ochs R.L., Stein T.W. Jr., Chan E.K.I., Ruutu M., Tan E.M.;  
 RT "cDNA cloning and characterization of a novel nucleolar protein.";  
 RL Mol. Biol. Cell 7:1015-1024(1996).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=20411121; PubMed=10952778;  
 RA Fossa A., Siebert R., Aasheim H.C., Maelandsmo G.M., Berner A.,  
 RA Fossa S.D., Paus E., Smeland E.B., Gaudernack G.;  
 RT "Identification of nucleolar protein No55 as a tumour-associated  
 autoantigen in patients with prostate cancer.";  
 RL Br. J. Cancer 83:743-749(2000).  
 RN [3]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=Brain;  
 RX MEDLINE=22388257; PubMed=12477932;  
 RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,  
 RA Klausner R.D., Collins P.S., Wagner L., Shenmen C.M., Schuler G.D.,  
 RA Altshul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,  
 RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,  
 RA Diatchenko L., Marusina K., Farmer A.A., Rubin G.W., Hong L.,  
 RA Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,  
 RA Brownstein M.J., Usdin T.B., Toshiyuki S., Carninci P., Prange C.,  
 RA Raha S.S., Locquellano N.A., Peters G.J., Abramson R.D., Mullahy S.J.,  
 RA Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,  
 RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,  
 RA Villalón D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,  
 RA Fahy J., Helton E., Kettman M., Madan A., Rodriguez S., Sanchez A.,  
 RA Whiting M., Madan A., Young A.C., Shevchenko Y., Soufard G.G.,  
 RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,  
 RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,  
 RA Butlerfield Y.S.N., Krzywinski M.I., Skalska U., Smallos D.E.,  
 RA Scherch A., Schein J.E., Jones S.J.M., Marra M.A.;  
 RT "Generation and initial analysis of more than 15,000 full-length  
 human and mouse cDNA sequences.";  
 RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).  
 CC -!- SUBCELLULAR LOCATION: Nuclear; nucleolar. Localized uniformly  
 throughout the granular component of the nucleolus and on the  
 surface of chromosomes during mitosis.  
 CC -!- SIMILARITY: BELONGS TO THE CRTAP / NO55 FAMILY.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 the European Bioinformatics Institute. There are no restrictions on its  
 use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 CC EMBL; U47621; AAC51792.1; -;  
 CC EMBL; AJ250583; CAC16786.1; -;  
 CC EMBL; BC001047; AAH01047.1; -;  
 CC EMBL; BC007942; AAH07942.1; -;  
 CC EMBL; BC011701; AAH11701.1; -;  
 CC GK; Q92791.1; -;  
 CC GO; GO:0000230; C:nuclear mitotic chromosome; TAS.  
 CC GO; GO:0005730; C:nucleolus; TAS.  
 CC GO; GO:0005716; C:synaptonemal complex; TAS.  
 CC GO; GO:0007130; P:synaptonemal complex formation; TAS.  
 CC InterPro; IPR008940; Prenyl\_trans.  
 KW Nuclear protein; Antigen.  
 FT DOMAIN 346 437 ASP/GLU-RICH (ACIDIC).  
 FT CONFLICT 186 186 Q -> R (IN REF. 2).

```
SQ SEQUENCE 437 AA; 50381 MW; 50C82FC9BB7274A CRC64;
Query Match 1.1%; Score 8; DB 1; Length 437;
Best Local Similarity 100.0%; Pred. No. 9.7;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 116 RAACLRRC 123
Db 112 RAACLRRC 119

RESULT 8
ID _UDPE NPVCD STANDARD; PRT; 493 AA.
AC O90158;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DE Ecdysteroid UDP-glucosyltransferase precursor (EC 2.4.1.-).
GN EGT OR UGT21A3.
OS Choristoneura fumiferana defective polyhedrosis virus (Cfdef).
OC Viruses; dsDNA viruses, no RNA stage; Baculoviridae;
OC Nucleopolyhedrovirus.
OX NCBI_TaxID=73480;
RN [1]
RP SEQUENCE FROM N.A.; PubMed=7595348;
RA Barrett J.W., Krell P.J., Arif B.M.;
RT "Characterization, sequencing and phylogeny of the ecdysteroid UDP-
glucosyltransferase gene from two distinct nuclear polyhedrosis
viruses isolated from Choristoneura fumiferana.";
RL J. Gen. Virol. 76:2447-2456(1995).
CC -!- FUNCTION: CATALYZES THE TRANSFER OF GLUCOSE FROM UDP-GLUCOSE TO
ECTYSTEROIDS WHICH ARE INSECT MOLTING HORMONES. EXPRESSION OF
EGT INTERFERES WITH NORMAL INSECT DEVELOPMENT AND BLOCK MOLTING
(BY SIMILARITY).
CC -!- DEVELOPMENTAL STAGE: PEAKED AROUND 12 HOURS P.I.
CC -!- SIMILARITY: Belongs to the UDP-glucosyltransferase family.
-----
This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.isb-sib.ch/announce/
or send an email to license@isb-sib.ch).
-----
EMBL: U10476; AAC09376.1; -.
DR InterPro: IPR002213; UDP_gluco_trans.
DR Pfam: PF00201; UDPGT; 1.
DR PROSITE: PS00375; UDPGT; 1.
KW Transferase; Glycosyltransferase; Signal.
FT SIGNAL 1 17 POTENTIAL.
FT CHAIN 18 493 ECDYSTEROID UDP-GLUCOSYLTRANSFERASE.
SQ SEQUENCE 493 AA; 54762 MW; 0C0BC9D60C3EF3A7 CRC64;

Query Match 1.1%; Score 8; DB 1; Length 493;
Best Local Similarity 100.0%; Pred. No. 11;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 8 LTTLLAV 15
Db 5 LTTLLAV 12

RESULT 9
ID _ONCKE STANDARD; PRT; 508 AA.
AC PA3371;
DT 28-FEB-2003 (Rel. 41, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DE Otolin-1 precursor.
DE Otolin-1 precursor.
```

```
GN OTOL.
OS Oncorhynchus keta (Chum salmon).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Actinopterygii; Neopterygii; Teleostei; Euteleostei;
OC Protacanthopterygii; Salmoniformes; Salmonidae; Oncorhynchus.
OX NCBI_TaxID=8018;
RN [1]
RP SEQUENCE FROM N.A.; SEQUENCE OF 26-40 AND 452-482, FUNCTION, AND
RP TISSUE SPECIFICITY.
RX MEDLINE=21845944; PubMed=11856329;
RA Murayama E., Takagi Y., Ohira T., Davis J.G., Greene M.I.,
RA Nagasawa H.;
RT "Fish otolith contains a unique structural protein, otolin-1.";
RL Eur. J. Biochem. 269:688-696(2002).
CC -!- FUNCTION: May be part of the internal framework of the otolith
CC where it may provide nucleation sites to facilitate
CC calcification.
CC -!- SUBCELLULAR LOCATION: Secreted.
CC -!- TISSUE SPECIFICITY: Selectively expressed in the sacculus where it
CC is localised to the otolith, the gelatinous layer of the otolithic
CC membrane, and part of the transitional epithelium.
CC -!- PTM: N-glycosylated.
CC -!- SIMILARITY: Contains 1 collagenous domain.
CC -!- SIMILARITY: Contains 1 Clq domain.
-----
This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.isb-sib.ch/announce/
or send an email to license@isb-sib.ch).
-----
EMBL: AB067770; BAB94561.1; -.
DR GO: GO:0005576; C:extracellular; IDA.
DR GO: GO:0045299; P:otolith mineralization; NAS.
DR InterPro: IPR001073; Clq.
DR InterPro: IPR008161; Clq_helix.
DR InterPro: IPR008160; Collagen.
DR InterPro: IPR008983; TNF_like.
DR Pfam: PF00386; Clq; 1.
DR Pfam: PF03391; Collagen; 4.
DR PRINTS: PR00007; COMPLENNTClQ.
DR PRODOM: PD000007; Clq_helix; 1.
DR PROSITE: PS01113; Clq; 1.
KW Collagen; Glycoprotein; Signal.
FT SIGNAL 1 25
FT CHAIN 26 508 OTOLIN-1.
FT DOMAIN 144 367 COLLAGEN-LIKE.
FT DOMAIN 369 505 ClQ.
FT CARBOHYD 121 121 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 416 416 N-LINKED (GLCNAC. .) (POTENTIAL).
SQ SEQUENCE 508 AA; 52138 MW; 90438799ACE3E9BE CRC64;

Query Match 1.1%; Score 8; DB 1; Length 508;
Best Local Similarity 100.0%; Pred. No. 11;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 9 LTTLLAV 16
Db 10 LTTLLAV 17

RESULT 10
ID IM13 ORYSA STANDARD; PRT; 84 AA.
AC O9XG5;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Mitochondrial import inner membrane translocase subunit Tim13.
GN TIM13.
OS Oryza sativa (Rice).
```

OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
OC Spermatophyta; Magnoliophyta; liliopsida; Poales; Poaceae;  
OC Ehrhartoideae; Oryzaceae; Oryza.  
OX NCBI TaxID=4530;

FT PROPEP 2 16 REMOVED IN MATURE FORM (BY SIMILARITY).  
FT CHAIN 17 66 TRANSFORMING GROWTH FACTOR ALPHA.  
FT PROPEP 67 121 REMOVED IN MATURE FORM (BY SIMILARITY).  
FT DOMAIN 2 75 EXTRACELLULAR (BY SIMILARITY).  
FT TRANSMEM 76 101 BY SIMILARITY.  
FT DOMAIN 20 60 EGF-LIKE (BY SIMILARITY).  
FT DISULFID 24 37 BY SIMILARITY.  
FT DISULFID 32 48 BY SIMILARITY.  
FT DISULFID 50 59 BY SIMILARITY.  
FT CARBOHYD 3 3 N-LINKED (GLCNAC... ) (POTENTIAL).  
FT NON TER 121 121  
SQ SEQUENCE 121 AA; 13182 MW; 463031DFBFC14816 CRC64;  
  
Query Match 1.0%; Score 7; DB 1; Length 121;  
Best Local Similarity 100.0%; Pred. No. 34; Indels 0; Gaps 0;  
Matches 7; Conservative 0; Mismatches 0;  
  
Qy 12 LLA VAA 18  
Db 64 LLA VAA 70  
|||||  
  
RESULT 13  
RL7 CHLITE STANDARD; PRT; 126 AA.  
AC O8KG16;  
DT 28-FEB-2003 (Rel. 41, Created)  
DT 28-FEB-2003 (Rel. 41, Last sequence update)  
DT 28-FEB-2003 (Rel. 41, Last annotation update)  
DE S58 ribosomal protein L7/L12.  
GN RPL7 OR CTO154.  
OS Chlorobium tepidum.  
OC Bacteria; Chlorobi; Chlorobia; Chlorobiales; Chlorobiaceae;  
OC Chlorobium.  
OX NCBI\_TaxID=1097;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=TLS / ATCC 49652 / DSM 12025;  
RX MEDLINE=22103685; PubMed=12093901;  
RA Eisen J.A., Nelson K.E., Paulsen I.T., Heidelberg J.F., Wu M.,  
RA Dodson R.J., Deboy R., Gwinn M.L., Nelson W.C., Kolonay J.L., Yang F.,  
RA Hickey E.K., Peterson J.D., Durkin A.S., Kolenay J.L., Parksey D.,  
RA Holt I., Umayam L.A., Mason T., Brenner M., Shea T.P., Parksey D.,  
RA Nierman W.C., Feldblyum T.V., Hansen C.L., Craven M.B., Radune D.,  
RA Vamathevan J., Khouri H., White O., Gruber T.M., Ketchum K.A.,  
RA Venter J.C., Tettelin H., Bryant D.A., Fraser C.M.;  
RT "The complete genome sequence of Chlorobium tepidum TLS, a  
photoautotrophic anaerobic, green-sulfur bacterium.";  
RL Proc. Natl. Acad. Sci. U.S.A. 99:9509-9514(2002).  
CC -!- FUNCTION: Seems to be the binding site for several of the factors  
involved in protein synthesis and appears to be essential for  
accurate translation (By similarity).  
CC -!- SIMILARITY: Belongs to the L12P family of ribosomal proteins.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
between the Swiss Institute of Bioinformatics and the EMBL outstation -  
the European Bioinformatics Institute. There are no restrictions on its  
use by non-profit institutions as long as its content is in no way  
modified and this statement is not removed. Usage by and for commercial  
entities requires a license agreement (see <http://www.isb-sib.ch/announce/>  
or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
CC EMBL; AE012794; AAM71402.1; -.  
DR TIGR; CT0154; -.  
DR HAMAP; MF\_00368; -; 1.  
DR InterPro; IPR008932; Ribos L12/7 olig.  
DR InterPro; IPR002026; Ribosomal L12.  
DR Pfam; PF00542; Ribosomal L12; 1.  
DR ProDom; PD001326; Ribosomal L12; 1.  
DR TrRFams; TIGR00855; L12; 1.  
KW Ribosomal protein; Complete proteome.  
SQ SEQUENCE 126 AA; 12996 MW; 31CFB56AC94FE054 CRC64;

Query Match 1.0%; Score 7; DB 1; Length 126;  
Best Local Similarity 100.0%; Pred. No. 35; Indels 0; Gaps 0;  
Matches 7; Conservative 0; Mismatches 0;  
  
Qy 428 IETLVEE 434  
Db 4 IETLVEE 10  
|||||  
  
RESULT 14  
TGFA SHEEP STANDARD; PRT; 133 AA.  
AC P98135;  
DT 01-FEB-1996 (Rel. 33, Created)  
DT 01-FEB-1996 (Rel. 33, Last sequence update)  
DT 10-OCT-2003 (Rel. 42, Last annotation update)  
DE Transforming growth factor alpha precursor (TGF-alpha) (EGF-like TGF)  
DE (ETGF) (TGF type 1) (Fragment).  
GN TGFA OR TGF-A.  
OS Ovis aries (Sheep).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;  
OC Bovidae; Caprinae; Ovis.  
OX NCBI\_TaxID=9940;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=Merino; TISSUE=Skin dorsal;  
RX MEDLINE=95268698; PubMed=7749621;  
RA Sutton R., Ward W.G., Raphael K.A., Cam G.R.;  
RT "Growth factor expression in skin during wool follicle development.";  
RL Comp. Biochem. Physiol. 110B:697-705(1995).  
CC -!- FUNCTION: TGF alpha is a mitogenic polypeptide that is able to  
bind to the EGF receptor and to act synergistically with TGF beta  
to promote anchorage-independent cell proliferation in soft agar.  
CC -!- SUBUNIT: Interacts with the PDZ domains of SDCBP and SNTAL. The  
interaction with SDCBP, is required for the targeting to the cell  
surface (By similarity).  
CC -!- SUBCELLULAR LOCATION: Type I membrane protein (precursor form);  
extracellular (mature form).  
CC -!- TISSUE SPECIFICITY: Skin.  
CC -!- DEVELOPMENTAL STAGE: Wool follicle development.  
CC -!- SIMILARITY: Contains 1 EGF-like domain.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
between the Swiss Institute of Bioinformatics and the EMBL outstation -  
the European Bioinformatics Institute. There are no restrictions on its  
use by non-profit institutions as long as its content is in no way  
modified and this statement is not removed. Usage by and for commercial  
entities requires a license agreement (see <http://www.isb-sib.ch/announce/>  
or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
CC EMBL; U36232; AAA53113.1; -.  
DR HSSP; P01135; 3TGF.  
DR InterPro; IPR001336; EGF 1.  
DR InterPro; IPR006209; EGF-like.  
DR InterPro; IPR006210; IEGF.  
DR PRINTS; PRO0009; EGF1TGF.  
DR SMART; SM00181; EGF; 1.  
DR PROSITE; PS00022; EGF 1; 1.  
DR PROSITE; PS01186; EGF 2; 1.  
DR PROSITE; PS00026; EGF 3; 1.  
KW Growth factor; Mitogen; Glycoprotein; EGF-like domain; Transmembrane;  
KW Signal.  
FT SIGNAL 1 23 POTENTIAL.  
FT PROPEP 24 38 REMOVED IN MATURE FORM.  
FT CHAIN 39 88 TRANSFORMING GROWTH FACTOR ALPHA.  
FT PROPEP 89 >133 REMOVED IN MATURE FORM.  
FT DOMAIN 24 97 EXTRACELLULAR (POTENTIAL).  
FT TRANSMEM 98 120 POTENTIAL.  
FT DOMAIN 121 >133 CYTOPLASMIC (POTENTIAL).  
FT DOMAIN 42 82 EGF-LIKE.  
FT DISULFID 46 59 BY SIMILARITY.  
FT DISULFID 54 70 BY SIMILARITY.

FT DISULFID 72 81 BY SIMILARITY.  
 FT CARBOHYD 25 25 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT NON TER 133 133  
 SQ SEQUENCE 133 AA; 14026 MW; F9F8E03BAA28AFB1 CRC64;

Query Match 1.0%; Score 7; DB 1; Length 133;  
 Best Local Similarity 100.0%; Pred.No. 37;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 12 LLAVVAA 18  
 DB 86 LLAVVAA 92

## RESULT 15

TGFA\_MOUSE  
 ID TGFA\_MOUSE STANDARD; PRT; 159 AA.  
 AC P48030;  
 DT 01-FEB-1996 (Rel. 33, Created)  
 DT 01-FEB-1996 (Rel. 33, Last sequence update)  
 DT 10-OCT-2003 (Rel. 42, Last annotation update)  
 DE Transforming growth factor alpha precursor (TGF-alpha) (EGF-like TGF)  
 DE (ETGF) (TGF type 1).  
 GN TGFA.  
 OS Mus musculus (Mouse).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 OX NCBI\_TaxID=10090;  
 RN [1]  
 RP SEQUENCE FROM N.A. PubMed=1420315;  
 RA MEDLINE=93041937; PubMed=1420315;  
 RA Vaughan T.J., Pascall J.C., Brown K.D.;  
 RA "Nucleotide sequence and tissue distribution of mouse transforming  
 growth factor-alpha."  
 RL Biochim. Biophys. Acta 1132:322-324(1992).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=129/SV; TISSUE=Brain;  
 RX MEDLINE=97031181; PubMed=8871107;  
 RA Berkowitz E.A., Serogy K.B., Schroeder J.A., Russell W.E.,  
 RA Evans E.P., Riedel R.F., Phillips H.K., Harrison C.A., Lee D.C.,  
 RA Luetke N.C.;  
 RA "Characterization of the mouse transforming growth factor alpha gene:  
 its expression during eyelid development and in waved 1 tissues."  
 RL Cell Growth Differ. 7:1271-1282(1996).  
 CC -!- FUNCTION: TGF alpha is a mitogenic polypeptide that is able to  
 bind to the EGF receptor and to act synergistically with TGF beta  
 to promote anchorage-independent cell proliferation in soft agar.  
 CC -!- SUBUNIT: Interacts with the PDZ domains of SDCBP and SNTAL. The  
 interaction with SDCBP is required for the targeting to the cell  
 surface (By similarity).  
 CC -!- SUBCELLULAR LOCATION: Type I membrane protein (precursor form);  
 extracellular (mature form).  
 CC -!- SIMILARITY: Contains 1 EGF-like domain.

-----  
 This SWISS-PROT entry is copyright. It is produced through a collaboration  
 between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 the European Bioinformatics Institute. There are no restrictions on its  
 use by non-profit institutions as long as its content is in no way  
 modified and this statement is not removed. Usage by and for commercial  
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 or send an email to [license@sb-sib.ch](mailto:license@sb-sib.ch)).  
 -----

EMBL; M92420; AAA04081.1; -  
 EMBL; U65016; AAB50554.1; -  
 EMBL; U64873; AAB50553.1; -  
 PIR; S27195; S27195.  
 PDB; 1GK5; 08-AUG-02.  
 DR MGI; 98724; Tgfa.  
 DR InterPro; IPR001336; EGF 1.  
 DR InterPro; IPR006209; EGF-like.  
 DR InterPro; IPR006210; IEGF.  
 DR Pfam; PF00008; EGF; 1.

DR PRINTS; PR00009; EGFTGF.  
 DR SMART; SM00181; EGF; 1.  
 DR PROSITE; PS00022; EGF\_1; 1.  
 DR PROSITE; PS01186; EGF\_2; 1.  
 DR PROSITE; PS00026; EGF\_3; 1.  
 KW Growth factor; Mitogen; Glycoprotein; EGF-like domain; Transmembrane;  
 KW Signal; Lipoprotein; Palmitate; 3D-structure.  
 FT SIGNAL 1 23  
 FT PROPEP 24 38  
 FT CHAIN 39 88  
 FT PROPEP 89 159  
 FT DOMAIN 24 97  
 FT TRANSMEM 98 123  
 FT DOMAIN 124 159  
 FT DOMAIN 44 83  
 FT DISULFID 46 59  
 FT DISULFID 54 70  
 FT DISULFID 72 81  
 FT LIPID 152 152  
 FT LIPID 153 153  
 FT CARBOHYD 25 25  
 SQ SEQUENCE 159 AA; 17018 MW; BE0AE8D9CE7DDF CRC64;

Query Match 1.0%; Score 7; DB 1; Length 159;  
 Best Local Similarity 100.0%; Pred.No. 43;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 12 LLAVVAA 18  
 DB 86 LLAVVAA 92

Search completed: July 18, 2004, 09:58:40  
 Job time : 18 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: July 18, 2004, 09:57:04 ; Search time 23 Seconds  
(without alignments)  
1652.031 Million cell updates/sec

Title: US-10-045-815-4  
Perfect score: 736  
Sequence: 1 MAVRALKLTTLILAVVAAS.....PPEPAQESUGSEKPKDEL 736

Scoring table: OLIGO  
Gapop 60.0 , Gapext 60.0

Searched: 389414 seqs, 51625971 residues

Word size : 0  
Total number of hits satisfying chosen parameters: 389414

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database : Issued Patents AA:\*  
1: /cgn2\_6/ptodata/2/iaa/5A COMB.pep.\*  
2: /cgn2\_6/ptodata/2/iaa/5B COMB.pep.\*  
3: /cgn2\_6/ptodata/2/iaa/6A COMB.pep.\*  
4: /cgn2\_6/ptodata/2/iaa/6B COMB.pep.\*  
5: /cgn2\_6/ptodata/2/iaa/PCTUS COMB.pep.\*  
6: /cgn2\_6/ptodata/2/iaa/backfiles1.pep.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match	Query Length	DB ID	Description
1	8	1.1	220	4	US-09-252-991A-19823
2	8	1.1	426	4	US-09-252-991A-20262
3	8	1.1	493	3	US-08-942-012B-28
4	8	1.1	613	4	US-09-489-039A-10059
5	7	1.0	50	1	US-08-208-008C-7
6	7	1.0	75	4	US-09-621-976-5127
7	7	1.0	82	1	US-07-847-743B-16
8	7	1.0	82	1	US-08-456-201-16
9	7	1.0	82	2	US-08-330-161-14
10	7	1.0	82	2	US-08-456-241-16
11	7	1.0	82	2	US-08-440-401-14
12	7	1.0	82	2	US-08-419-878B-14
13	7	1.0	82	4	US-09-173-480-14
14	7	1.0	82	5	PCT-US92-04295A-16
15	7	1.0	87	2	US-07-885-089B-12
16	7	1.0	115	3	US-08-545-089A-122
17	7	1.0	124	4	US-09-489-039A-11363
18	7	1.0	125	4	US-09-489-039A-10706
19	7	1.0	159	4	US-09-227-853A-14
20	7	1.0	171	3	US-08-611-587-5
21	7	1.0	177	4	US-09-252-991A-29834
22	7	1.0	184	4	US-09-252-991A-32941
23	7	1.0	185	4	US-09-540-236-3570
24	7	1.0	189	4	US-09-252-991A-18839
25	7	1.0	202	4	US-09-543-681A-7996
26	7	1.0	210	3	US-08-611-587-4
27	7	1.0	211	4	US-09-252-991A-26393

28 7 1.0 219 4 US-08-311-731A-305 Sequence 305, App  
29 7 1.0 243 4 US-09-252-991A-19011 Sequence 19011, A  
30 7 1.0 274 4 US-09-489-039A-13000 Sequence 13000, A  
31 7 1.0 290 4 US-09-134-001C-3401 Sequence 3401, Ap  
32 7 1.0 297 4 US-09-584-568C-7 Sequence 7, Appli  
33 7 1.0 302 4 US-09-252-991A-30761 Sequence 30761, A  
34 7 1.0 316 4 US-09-711-184-378 Sequence 378, App  
35 7 1.0 325 4 US-09-252-991A-18010 Sequence 18010, A  
36 7 1.0 345 1 US-08-843-993-1 Sequence 1, Appli  
37 7 1.0 345 3 US-09-059-520A-1 Sequence 1, Appli  
38 7 1.0 345 3 US-09-334-275-1 Sequence 1, Appli  
39 7 1.0 351 4 US-09-540-236-3819 Sequence 17906, A  
40 7 1.0 352 4 US-09-252-991A-17906 Sequence 6547, Ap  
41 7 1.0 352 4 US-09-328-352-6547 Sequence 28480, A  
42 7 1.0 357 4 US-09-252-991A-28480 Sequence 12, Appl  
43 7 1.0 364 4 US-09-242-859A-12 Sequence 37, Appl  
44 7 1.0 387 3 US-08-968-563-37 Sequence 37, Appl  
45 7 1.0 387 3 US-08-969-683A-37 Sequence 37, Appl

ALIGNMENTS

RESULT 1  
US-09-252-991A-19823  
; Sequence 19823, Application US/09252991A  
; Patent No. 6551795  
; GENERAL INFORMATION:  
; APPLICANT: Marc J. Rubenfield et al.  
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS  
; FILE REFERENCE: 107196.136  
; CURRENT APPLICATION NUMBER: US/09/252,991A  
; CURRENT FILING DATE: 1999-02-18  
; PRIOR APPLICATION NUMBER: US 60/074,788  
; PRIOR FILING DATE: 1998-02-18  
; PRIOR APPLICATION NUMBER: US 60/094,190  
; PRIOR FILING DATE: 1998-07-27  
; NUMBER OF SEQ ID NOS: 33142  
; SEQ ID NO 19823  
; LENGTH: 220  
; TYPE: PRT  
; ORGANISM: Pseudomonas aeruginosa  
US-09-252-991A-19823

Query Match 1.1%; Score 8; DB 4; Length 220;  
Best Local Similarity 100.0%; Pred. No. 17; Mismatches 0; Indels 0; Gaps 0;  
Matches 8; Conservative 0;

OY 96 AQASGAGA 103  
|||||  
Db 7 AQASGAGA 14

RESULT 2  
US-09-252-991A-20262  
; Sequence 20262, Application US/09252991A  
; Patent No. 6551795  
; GENERAL INFORMATION:  
; APPLICANT: Marc J. Rubenfield et al.  
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS  
; FILE REFERENCE: 107196.136  
; CURRENT APPLICATION NUMBER: US/09/252,991A  
; CURRENT FILING DATE: 1999-02-18  
; PRIOR APPLICATION NUMBER: US 60/074,788  
; PRIOR FILING DATE: 1998-02-18  
; PRIOR APPLICATION NUMBER: US 60/094,190  
; PRIOR FILING DATE: 1998-07-27  
; NUMBER OF SEQ ID NOS: 33142  
; SEQ ID NO 20262  
; LENGTH: 426  
; TYPE: PRT



/ ORGANISM: Pseudomonas aeruginosa  
US-09-252-991A-20262

Query Match 1.1%; Score 8; DB 4; Length 426;  
Best Local Similarity 100.0%; Pred. No. 30;  
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 115 RRAACLR 122  
| | | | |  
Db 206 RRAACLR 213

RESULT 3  
US-08-942-012B-28

/ Sequence 28, Application US/08942012B  
/ Patent No. 6235278  
/ GENERAL INFORMATION:  
/ APPLICANT: Miller, Lois K.  
/ APPLICANT: Lu, Albert  
/ APPLICANT: Dierks, Peter  
/ APPLICANT: Black, Bruce  
/ TITLE OF INVENTION: Biological Insect Control Agents Expressing  
/ TITLE OF INVENTION: Insect-Specific Toxin Genes, Methods and Compositions  
/ FILE REFERENCE: 28-96a  
/ CURRENT APPLICATION NUMBER: US/08/942,012B  
/ CURRENT FILING DATE: 1997-10-01  
/ PRIOR APPLICATION NUMBER: 08/729,606  
/ PRIOR FILING DATE: 2000-10-01  
/ NUMBER OF SEQ ID NOS: 33  
/ SOFTWARE: Patent In Ver. 2.0  
/ SEQ ID NO 28  
/ LENGTH: 493  
/ TYPE: PRT  
/ ORGANISM: Cf defective nucleopolyhedrosis virus  
US-08-942-012B-28

Query Match 1.1%; Score 8; DB 3; Length 493;  
Best Local Similarity 100.0%; Pred. No. 34;  
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 8 LLTLLAV 15  
| | | | |  
Db 5 LLTLLAV 12

RESULT 4  
US-09-489-039A-10059

/ Sequence 10059, Application US/09489039A  
/ Patent No. 6610836  
/ GENERAL INFORMATION:  
/ APPLICANT: Gary Breton et. al  
/ TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO KLEBSIELLA  
/ TITLE OF INVENTION: PNEUMONIAE FOR DIAGNOSTICS AND THERAPEUTICS  
/ FILE REFERENCE: 2709.2004001  
/ CURRENT APPLICATION NUMBER: US/09/489,039A  
/ CURRENT FILING DATE: 2000-01-27  
/ PRIOR APPLICATION NUMBER: US 60/117,747  
/ PRIOR FILING DATE: 1999-01-29  
/ NUMBER OF SEQ ID NOS: 14342  
/ SEQ ID NO 10059  
/ LENGTH: 613  
/ TYPE: PRT  
/ ORGANISM: Klebsiella pneumoniae  
US-09-489-039A-10059

Query Match 1.1%; Score 8; DB 4; Length 613;  
Best Local Similarity 100.0%; Pred. No. 41;  
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 20 SOAEVESE 27  
| | | | |  
Db 407 SOAEVESE 414

RESULT 5  
US-08-208-008C-7

/ Sequence 7, Application US/08208008C  
/ Patent No. 5633147  
/ GENERAL INFORMATION:  
/ APPLICANT: MEISSNER, ET AL.  
/ TITLE OF INVENTION: Transforming Growth Factor  
/ TITLE OF INVENTION: Alpha - H1  
/ NUMBER OF SEQUENCES: 14  
/ CORRESPONDENCE ADDRESS:  
/ ADDRESSEE: CARELLA, BYRNE, BAIN, GILFILLAN,  
/ ADDRESSEE: CECCHI, STEWART & OLSTEIN  
/ STREET: 6 BECKER FARM ROAD  
/ CITY: ROSELAND  
/ STATE: NEW JERSEY  
/ COUNTRY: USA  
/ ZIP: 07068  
/ COMPUTER READABLE FORM:  
/ MEDIUM TYPE: 3.5 INCH DISKETTE  
/ COMPUTER: IBM PS/2  
/ OPERATING SYSTEM: MS-DOS  
/ SOFTWARE: WORD PERFECT 5.1  
/ CURRENT APPLICATION DATA:  
/ APPLICATION NUMBER: US/08/208,008C  
/ FILING DATE: March 8, 1994  
/ CLASSIFICATION: 435  
/ PRIOR APPLICATION DATA:  
/ APPLICATION NUMBER:  
/ FILING DATE:  
/ ATTORNEY/AGENT INFORMATION:  
/ NAME: MULLINS, J.G.  
/ REGISTRATION NUMBER: 33,073  
/ REFERENCE/DOCKET NUMBER: 325800-98  
/ TELECOMMUNICATION INFORMATION:  
/ TELEPHONE: 201-994-1700  
/ TELEFAX: 201-994-1744  
/ INFORMATION FOR SEQ ID NO: 7:  
/ SEQUENCE CHARACTERISTICS:  
/ LENGTH: 50 AMINO ACIDS  
/ TYPE: AMINO ACID  
/ STRANDEDNESS:  
/ TOPOLOGY: LINEAR  
/ MOLECULE TYPE: PROTEIN  
US-08-208-008C-7

Query Match 1.0%; Score 7; DB 1; Length 50;  
Best Local Similarity 100.0%; Pred. No. 42;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 12 LLAVVAA 18  
| | | | |  
Db 43 LLAVVAA 49

RESULT 6

US-09-621-976-5127  
/ Sequence 5127, Application US/09621976  
/ Patent No. 6639063  
/ GENERAL INFORMATION:  
/ APPLICANT: Dumas Milne Edwards, J.B.  
/ APPLICANT: Jobert, S.  
/ APPLICANT: Giordano, J.Y.  
/ TITLE OF INVENTION: ESTs and Encoded Human Proteins.  
/ FILE REFERENCE: GENSET.054PR2  
/ CURRENT APPLICATION NUMBER: US/09/621,976  
/ CURRENT FILING DATE: 2000-07-21  
/ NUMBER OF SEQ ID NOS: 19335  
/ SOFTWARE: Patent.pm  
/ SEQ ID NO 5127  
/ LENGTH: 75  
/ TYPE: PRT  
/ ORGANISM: Homo sapiens

FEATURE:  
NAME/KEY: UNSURE  
LOCATION: 62  
OTHER INFORMATION: Xaa = Gly,Ser  
NAME/KEY: UNSURE  
LOCATION: 69  
OTHER INFORMATION: Xaa = Ile,Thr  
US-09-621-976-5127

Query Match 1.0%; Score 7; DB 4; Length 75;  
Best Local Similarity 100.0%; Pred. No. 60;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 176 HMEQOON 182  
Db 54 HMEQOON 60

RESULT 7  
US-07-847-743B-16  
Sequence 16, Application US/07847743B  
Patent No. 5367060  
GENERAL INFORMATION:  
APPLICANT: Genentech, Inc.  
TITLE OF INVENTION: Structure, Production and Use of  
TITLE OF INVENTION: Heregulin  
NUMBER OF SEQUENCES: 30  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Genentech, Inc.  
STREET: 460 Point San Bruno Blvd  
CITY: South San Francisco  
STATE: California  
COUNTRY: USA  
ZIP: 94080  
COMPUTER READABLE FORM:  
MEDIUM TYPE: 5.25 inch, 360 Kb floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: patin (Genentech)  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/07/847,743B  
FILING DATE: 19920306  
CLASSIFICATION: 424  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 07/705256  
FILING DATE: 24-May-1991  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 07/765212  
FILING DATE: 25-Sep-1991  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 07/790801  
FILING DATE: 08-No. 5367060-1991  
ATTORNEY/AGENT INFORMATION:  
NAME: Hensley, Max D.  
REGISTRATION NUMBER: 27,043  
REFERENCE/DOCKET NUMBER: 712P3  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 415/266-1489  
TELEFAX: 415/952-9881  
TELEX: 910/371-7168  
INFORMATION FOR SEQ ID NO: 16:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 82 amino acids  
TYPE: AMINO ACID  
TOPOLOGY: linear

US-07-847-743B-16

Query Match 1.0%; Score 7; DB 1; Length 82;  
Best Local Similarity 100.0%; Pred. No. 65;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 12 LLAVVAA 18  
Db 43 LLAVVAA 49

Db 43 LLAVVAA 49.

RESULT 8  
US-08-456-201-16  
Sequence 16, Application US/08456201  
Patent No. 5641869  
GENERAL INFORMATION:  
APPLICANT: Vandlen, Richard L.,  
TITLE OF INVENTION: Structure, Production and Use of  
TITLE OF INVENTION: Heregulin 2 Ligands  
NUMBER OF SEQUENCES: 30  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Genentech, Inc.  
STREET: 460 Point San Bruno Blvd  
CITY: South San Francisco  
STATE: California  
COUNTRY: USA  
ZIP: 94080  
COMPUTER READABLE FORM:  
MEDIUM TYPE: 5.25 inch, 360 Kb floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patin (Genentech)  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/456,201  
FILING DATE:  
CLASSIFICATION: 530  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US/08/126,145  
FILING DATE:  
APPLICATION NUMBER: 07/880,917  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 07/705256  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 07/765212  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 07/790801  
ATTORNEY/AGENT INFORMATION:  
NAME: Hensley, Max D.  
REGISTRATION NUMBER: 27,034  
REFERENCE/DOCKET NUMBER: 712P4  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 415/266-1489  
TELEFAX: 415/952-9881  
TELEX: 910/371-7168  
INFORMATION FOR SEQ ID NO: 16:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 82 amino acids  
TYPE: amino acid  
TOPOLOGY: linear

US-08-456-201-16

Query Match 1.0%; Score 7; DB 1; Length 82;  
Best Local Similarity 100.0%; Pred. No. 65;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 12 LLAVVAA 18  
Db 43 LLAVVAA 49

RESULT 9  
US-08-330-161-14  
Sequence 14, Application US/08330161  
Patent No. 5834229  
GENERAL INFORMATION:  
APPLICANT: Vandlen, Richard  
APPLICANT: Holmes, William  
TITLE OF INVENTION: Structure, Production and Use of Heregulin 2 Ligands  
NUMBER OF SEQUENCES: 17  
CORRESPONDENCE ADDRESS:

ADDRESSEE: Genentech, Inc.  
 STREET: 460 Point San Bruno Blvd  
 CITY: South San Francisco  
 STATE: California  
 COUNTRY: USA  
 ZIP: 94080  
 COMPUTER READABLE FORM:  
 MEDIUM TYPE: 5.25 inch, 360 Kb floppy disk  
 COMPUTER: IBM PC compatible  
 OPERATING SYSTEM: PC-DOS/MS-DOS  
 SOFTWARE: patin (Genentech)  
 CURRENT APPLICATION DATA:  
 APPLICATION NUMBER: US/08/330,161  
 FILING DATE:  
 CLASSIFICATION: 435  
 PRIOR APPLICATION DATA:  
 APPLICATION NUMBER: 08/035430  
 FILING DATE: 22-MAR-1993  
 APPLICATION NUMBER: 07/705256  
 FILING DATE: 24-MAY-1991  
 ATTORNEY/AGENT INFORMATION:  
 NAME: Adler, Carolyn R.  
 REGISTRATION NUMBER: 32,324  
 REFERENCE/DOCKET NUMBER: 712C1  
 TELECOMMUNICATION INFORMATION:  
 TELEPHONE: 415/225-2614  
 TELEFAX: 415/952-9881  
 TELEX: 910/371-7168  
 INFORMATION FOR SEQ ID NO: 14:  
 SEQUENCE CHARACTERISTICS:  
 LENGTH: 82 amino acids  
 TYPE: amino acid  
 TOPOLOGY: linear  
 US-08-330-161-14

Query Match 1.0%; Score 7; DB 2; Length 82;  
 Best Local Similarity 100.0%; Pred. No. 65;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 12 LLAVVAA 18  
 DB 43 LLAVVAA 49

RESULT 10  
 US-08-456-241-16  
 ; Sequence 16, Application US/08456241  
 ; Patent No. 5840525  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Vandlen, Richard L.  
 ; APPLICANT: Holmes, William E.  
 ; TITLE OF INVENTION: STRUCTURE, PRODUCTION AND USE OF  
 ; TITLE OF INVENTION: HEREGULIN  
 ; NUMBER OF SEQUENCES: 30  
 ; CORRESPONDENCE ADDRESS:  
 ; ADDRESSEE: Genentech, Inc.  
 ; STREET: 460 Point San Bruno Blvd  
 ; CITY: South San Francisco  
 ; STATE: California  
 ; COUNTRY: USA  
 ; ZIP: 94080  
 ; COMPUTER READABLE FORM:  
 ; MEDIUM TYPE: 5.25 inch, 360 Kb floppy disk  
 ; COMPUTER: IBM PC compatible  
 ; OPERATING SYSTEM: PC-DOS/MS-DOS  
 ; SOFTWARE: patin (Genentech)  
 ; CURRENT APPLICATION DATA:  
 ; APPLICATION NUMBER: US/08/456,241  
 ; FILING DATE: 31-MAY-1995  
 ; CLASSIFICATION: 435  
 ; PRIOR APPLICATION DATA:  
 ; APPLICATION NUMBER: 08/126145  
 ; FILING DATE: 23-SEP-1993

PRIOR APPLICATION DATA:  
 APPLICATION NUMBER: 07/880917  
 FILING DATE: 11-MAY-1992  
 PRIOR APPLICATION DATA:  
 APPLICATION NUMBER: 07/847743  
 FILING DATE: 06-MAR-1992  
 PRIOR APPLICATION DATA:  
 APPLICATION NUMBER: 07/790801  
 FILING DATE: 08-NOV-1991  
 PRIOR APPLICATION DATA:  
 APPLICATION NUMBER: 07/765212  
 FILING DATE: 25-SEP-1991  
 PRIOR APPLICATION DATA:  
 APPLICATION NUMBER: 07/705256  
 FILING DATE: 24-MAY-1991  
 ATTORNEY/AGENT INFORMATION:  
 NAME: Lee, Wendy M.  
 REGISTRATION NUMBER: 00,000  
 REFERENCE/DOCKET NUMBER: 712P4C1D1  
 TELECOMMUNICATION INFORMATION:  
 TELEPHONE: 415/225-1994  
 TELEFAX: 415/952-9881  
 TELEX: 910/371-7168  
 INFORMATION FOR SEQ ID NO: 16:  
 SEQUENCE CHARACTERISTICS:  
 LENGTH: 82 amino acids  
 TYPE: amino acid  
 TOPOLOGY: linear  
 US-08-456-241-16

Query Match 1.0%; Score 7; DB 2; Length 82;  
 Best Local Similarity 100.0%; Pred. No. 65;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 12 LLAVVAA 18  
 DB 43 LLAVVAA 49

RESULT 11  
 US-08-440-401-14  
 ; Sequence 14, Application US/08440401  
 ; Patent No. 5856110  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Vandlen, Richard L.  
 ; APPLICANT: Holmes, William E.  
 ; TITLE OF INVENTION: Structure, Production and Use of  
 ; TITLE OF INVENTION: Heregulin 2 Ligands  
 ; NUMBER OF SEQUENCES: 17  
 ; CORRESPONDENCE ADDRESS:  
 ; ADDRESSEE: Genentech, Inc.  
 ; STREET: 460 Point San Bruno Blvd  
 ; CITY: South San Francisco  
 ; STATE: California  
 ; COUNTRY: USA  
 ; ZIP: 94080  
 ; COMPUTER READABLE FORM:  
 ; MEDIUM TYPE: 5.25 inch, 360 Kb floppy disk  
 ; COMPUTER: IBM PC compatible  
 ; OPERATING SYSTEM: PC-DOS/MS-DOS  
 ; SOFTWARE: patin (Genentech)  
 ; CURRENT APPLICATION DATA:  
 ; APPLICATION NUMBER: US/08/440,401  
 ; FILING DATE: 12-MAY-1995  
 ; CLASSIFICATION: 435  
 ; PRIOR APPLICATION DATA:  
 ; APPLICATION NUMBER: 08/330161  
 ; FILING DATE: 25-OCT-1994  
 ; PRIOR APPLICATION DATA:  
 ; APPLICATION NUMBER: 08/035430  
 ; FILING DATE: 22-MAR-1993  
 ; PRIOR APPLICATION DATA:  
 ; APPLICATION NUMBER: 07/705256

FILING DATE: 24-MAY-1991  
ATTORNEY/AGENT INFORMATION:  
NAME: Lee, Wendy M.  
REGISTRATION NUMBER: 00,000  
REFERENCE/DOCKET NUMBER: 712C3  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 415/225-1994  
TELEFAX: 415/952-9881  
TELEX: 910/371-7168  
INFORMATION FOR SEQ ID NO: 14:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 82 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
US-08-440-401-14

Query Match 1.0%; Score 7; DB 2; Length 82;  
Best Local Similarity 100.0%; Pred. No. 65;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 12 LLAVVAA 18  
Db 43 LLAVVAA 49

RESULT 13  
US-09-173-480-14  
Sequence 14, Application US/09173480  
Patent No. 6399746  
GENERAL INFORMATION:  
APPLICANT: Vandlen, Richard L.  
Holmes, William E.  
TITLE OF INVENTION: Structure, Production and Use of Heregulin 2 Ligands  
NUMBER OF SEQUENCES: 17  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Genentech, Inc.  
STREET: 460 Point San Bruno Blvd  
CITY: South San Francisco  
STATE: California  
COUNTRY: USA  
ZIP: 94080  
COMPUTER READABLE FORM:  
MEDIUM TYPE: 5.25 inch, 360 Kb floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patin (Genentech)  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/173,480  
FILING DATE: 14-Oct-1998  
CLASSIFICATION: <Unknown>  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 08/440,401  
FILING DATE: <Unknown>  
APPLICATION NUMBER: 08/035430  
FILING DATE: 22-MAR-1993  
APPLICATION NUMBER: 07/705256  
FILING DATE: 24-MAY-1991  
ATTORNEY/AGENT INFORMATION:  
NAME: Lee, Wendy M.  
REGISTRATION NUMBER: 00,000  
REFERENCE/DOCKET NUMBER: 712C3  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 415/225-1994  
TELEFAX: 415/952-9881  
TELEX: 910/371-7168  
INFORMATION FOR SEQ ID NO: 14:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 82 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
SEQUENCE DESCRIPTION: SEQ ID NO: 14:  
US-09-173-480-14

Query Match 1.0%; Score 7; DB 4; Length 82;  
Best Local Similarity 100.0%; Pred. No. 65;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 12 LLAVVAA 18  
Db 43 LLAVVAA 49

RESULT 14  
PCT-US92-04295A-16  
Sequence 16, Application PC/TUS9204295A  
GENERAL INFORMATION:  
APPLICANT: Genentech, Inc.  
TITLE OF INVENTION: Structure, Production and Use of

FILING DATE: 24-MAY-1991  
ATTORNEY/AGENT INFORMATION:  
NAME: Lee, Wendy M.  
REGISTRATION NUMBER: 00,000  
REFERENCE/DOCKET NUMBER: 712C3  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 415/225-1994  
TELEFAX: 415/952-9881  
TELEX: 910/371-7168  
INFORMATION FOR SEQ ID NO: 14:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 82 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
US-08-440-401-14

Query Match 1.0%; Score 7; DB 2; Length 82;  
Best Local Similarity 100.0%; Pred. No. 65;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 12 LLAVVAA 18  
Db 43 LLAVVAA 49

RESULT 12  
US-08-419-878B-14  
Sequence 14, Application US/08419878B  
Patent No. 5859206  
GENERAL INFORMATION:  
APPLICANT: Vandlen, Richard L.  
Holmes, William E.  
TITLE OF INVENTION: Antibodies Specific For Heregulin 2-alpha  
NUMBER OF SEQUENCES: 18  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Genentech, Inc.  
STREET: 1 DNA Way  
CITY: South San Francisco  
STATE: California  
COUNTRY: USA  
ZIP: 94080  
COMPUTER READABLE FORM:  
MEDIUM TYPE: 3.5 inch, 1.44 Mb floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: WinPatin (Genentech)  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/419,878B  
FILING DATE: 11-Apr-1995  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 08/330161  
FILING DATE: 25-OCT-1994  
APPLICATION NUMBER: 08/035430  
FILING DATE: 22-MAR-1993  
ATTORNEY/AGENT INFORMATION:  
NAME: Lee, Wendy M.  
REGISTRATION NUMBER: 40,378  
REFERENCE/DOCKET NUMBER: P0712C2D1  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 650/225-1994  
TELEFAX: 650/952-9881  
INFORMATION FOR SEQ ID NO: 14:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 82 amino acids  
TYPE: Amino Acid  
TOPOLOGY: Linear  
US-08-419-878B-14

;; TITLE OF INVENTION: Heregulin  
;; NUMBER OF SEQUENCES: 30  
;; CORRESPONDENCE ADDRESS:  
;; ADDRESSEE: Genentech, Inc.  
;; STREET: 460 Point San Bruno Blvd  
;; CITY: South San Francisco  
;; STATE: California  
;; COUNTRY: USA  
;; ZIP: 94080  
;;  
;; COMPUTER READABLE FORM:  
;; MEDIUM TYPE: 5.25 inch, 360 Kb floppy disk  
;; COMPUTER: IBM PC compatible  
;; OPERATING SYSTEM: PC-DOS/MS-DOS  
;; SOFTWARE: patin (Genentech)  
;; CURRENT APPLICATION DATA:  
;; APPLICATION NUMBER: PCT/US92/04295A  
;; FILING DATE: 19920521  
;; CLASSIFICATION:  
;; PRIOR APPLICATION DATA:  
;; APPLICATION NUMBER: 07/705256  
;; FILING DATE: 24-MAY-1991  
;; PRIOR APPLICATION DATA:  
;; APPLICATION NUMBER: 07/765212  
;; FILING DATE: 25-SEP-1991  
;; PRIOR APPLICATION DATA:  
;; APPLICATION NUMBER: 07/790801  
;; FILING DATE: 08-NOV-1991  
;; PRIOR APPLICATION DATA:  
;; APPLICATION NUMBER: 07/847743  
;; FILING DATE: 06-MAR-1992  
;; ATTORNEY/AGENT INFORMATION:  
;; NAME: Hensley, Max D.  
;; REGISTRATION NUMBER: 27,043  
;; REFERENCE/DOCKET NUMBER: 712P4  
;; TELECOMMUNICATION INFORMATION:  
;; TELEPHONE: 415/225-1994  
;; TELEFAX: 415/352-9881  
;; TELEX: 910/371-7168  
;; INFORMATION FOR SEQ ID NO: 16:  
;; SEQUENCE CHARACTERISTICS:  
;; LENGTH: 82 amino acids  
;; TYPE: AMINO ACID  
;; TOPOLOGY: linear  
;;  
PCT-US92-04295A-16

Query Match 1.0%; Score 7; DB 5; Length 82;  
Best Local Similarity 100.0%; Pred. No. 65;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 12 LLAVVAA 18  
Db 43 LLAVVAA 49

RESULT 15  
US-07-885-089B-12  
; Sequence 12, Application US/07885089B  
; Parent No. 5830595  
; GENERAL INFORMATION:  
; APPLICANT: Shoyab, Mohammed  
; APPLICANT: McDonald, Vicki L.  
; APPLICANT: Bradley, James G.  
; APPLICANT: Plowman, Gregory D.  
; TITLE OF INVENTION: AMPHIREGULINS: A FAMILY OF  
; TITLE OF INVENTION: HEPARIN-BINDING EPITHELIAL CELL GROWTH FACTORS  
; NUMBER OF SEQUENCES: 37  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Pennie & Edmonds LLP  
; STREET: 1155 Avenue of the Americas  
; CITY: New York  
; STATE: New York  
; COUNTRY: USA  
; ZIP: 10036

;; COMPUTER READABLE FORM:  
;; MEDIUM TYPE: Floppy disk  
;; COMPUTER: IBM PC compatible  
;; OPERATING SYSTEM: PC-DOS/MS-DOS  
;; SOFTWARE: Patent in Release #1.0, Version #1.30  
;; CURRENT APPLICATION DATA:  
;; APPLICATION NUMBER: US/07/885,089B  
;; FILING DATE: 18-MAY-1992  
;; CLASSIFICATION: 530  
;; ATTORNEY/AGENT INFORMATION:  
;; NAME: Coruzzi, Laura A.  
;; REGISTRATION NUMBER: 30,742  
;; REFERENCE/DOCKET NUMBER: 5624-174  
;; TELECOMMUNICATION INFORMATION:  
;; TELEPHONE: 212-790-9090  
;; TELEFAX: 212-869-9741  
;; TELEX: 66141 PENNIE  
;; INFORMATION FOR SEQ ID NO: 12:  
;; SEQUENCE CHARACTERISTICS:  
;; LENGTH: 87 amino acids  
;; TYPE: amino acid  
;; STRANDEDNESS: single  
;; TOPOLOGY: linear  
;; MOLECULE TYPE: peptide  
;;  
US-07-885-089B-12

Query Match 1.0%; Score 7; DB 2; Length 87;  
Best Local Similarity 100.0%; Pred. No. 69;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 12 LLAVVAA 18  
Db 41 LLAVVAA 47

Search completed: July 18, 2004, 10:00:58  
Job time : 24 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: July 18, 2004, 09:58:14 ; Search time 53 Seconds  
(without alignments)  
4340.465 Million cell updates/sec

Title: US-10-045-815-4

Perfect score: 736

Sequence: 1 MAVRALKLLTLLAVVAAS.....PPPAQESLSGSBSKPKDEL 736

Scoring table: OIIGO

Gapop 60.0 , Gapext 60.0

Searched: 1285345 seqs, 31256633 residues

Word size : 0

Total number of hits satisfying chosen parameters: 1285345

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database : Published Applications AA:

- 1: /cgn2\_6/prodata/2/pubpaa/US07\_PUBCOMB.pep.\*
- 2: /cgn2\_6/prodata/2/pubpaa/PCT\_NEW\_PUB.pep.\*
- 3: /cgn2\_6/prodata/2/pubpaa/US06\_NEW\_PUB.pep.\*
- 4: /cgn2\_6/prodata/2/pubpaa/US06\_PUBCOMB.pep.\*
- 5: /cgn2\_6/prodata/2/pubpaa/US07\_NEW\_PUB.pep.\*
- 6: /cgn2\_6/prodata/2/pubpaa/PCTUS\_PUBCOMB.pep.\*
- 7: /cgn2\_6/prodata/2/pubpaa/US08\_NEW\_PUB.pep.\*
- 8: /cgn2\_6/prodata/2/pubpaa/US08\_PUBCOMB.pep.\*
- 9: /cgn2\_6/prodata/2/pubpaa/US09A\_PUBCOMB.pep.\*
- 10: /cgn2\_6/prodata/2/pubpaa/US09B\_PUBCOMB.pep.\*
- 11: /cgn2\_6/prodata/2/pubpaa/US09C\_PUBCOMB.pep.\*
- 12: /cgn2\_6/prodata/2/pubpaa/US09\_NEW\_PUB.pep.\*
- 13: /cgn2\_6/prodata/2/pubpaa/US10A\_PUBCOMB.pep.\*
- 14: /cgn2\_6/prodata/2/pubpaa/US10B\_PUBCOMB.pep.\*
- 15: /cgn2\_6/prodata/2/pubpaa/US10C\_PUBCOMB.pep.\*
- 16: /cgn2\_6/prodata/2/pubpaa/US10\_NEW\_PUB.pep.\*
- 17: /cgn2\_6/prodata/2/pubpaa/US60\_NEW\_PUB.pep.\*
- 18: /cgn2\_6/prodata/2/pubpaa/US60\_PUBCOMB.pep.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	736	100.0	736	13	US-10-045-815-4
2	433	58.8	736	12	US-10-257-174-26
3	433	58.8	736	12	US-10-312-352-33
4	360	48.9	363	13	US-10-045-815-2
5	275	37.4	708	12	US-10-257-174-25
6	219	29.8	359	11	US-09-833-245-1251
7	101	13.7	173	11	US-09-833-245-1250
8	91	12.4	747	13	US-10-045-815-6
9	82	11.1	542	13	US-10-045-815-8
10	13	1.8	219	11	US-09-764-875-815
11	13	1.8	267	9	US-09-764-868-736
12	13	1.8	287	11	US-09-764-875-1106
13	13	1.8	282	11	US-10-012-697-1493
14	13	1.8	708	15	US-10-094-749-2560
15	8	1.1	10	10	US-09-572-404B-1740

#### ALIGNMENTS

##### RESULT 1

US-10-045-815-4  
; Sequence 4, Application US/10045815  
; Publication No. US20020160498A1  
; GENERAL INFORMATION:  
; APPLICANT: Wadhwa, Renu  
; APPLICANT: Sugihara, Takashi  
; APPLICANT: Ohide, Akiko  
; TITLE OF INVENTION: TUMOR SUPPRESSOR GENE  
; FILE REFERENCE: 06501-091001  
; CURRENT APPLICATION NUMBER: US/10/045,815  
; CURRENT FILING DATE: 2001-10-26  
; PRIOR APPLICATION NUMBER: PCI/JPO00/02731  
; PRIOR FILING DATE: 2000-04-26  
; PRIOR APPLICATION NUMBER: JP 11/118806  
; PRIOR FILING DATE: 1999-04-26  
; NUMBER OF SEQ ID NOS: 15  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 4  
; LENGTH: 736  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
US-10-045-815-4

Query Match 100.0%; Score 736; DB 13; Length 736;  
Best Local Similarity 100.0%; Pred. No. 0;  
Matches 736; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1 MAVRALKLLTLLAVVAASQAQVESEAGWGMVTPDLLFAEGTAAYARGDWPVGLSMER 60  
Db 1 MAVRALKLLTLLAVVAASQAQVESEAGWGMVTPDLLFAEGTAAYARGDWPVGLSMER 60  
QY 61 ALRSPAALRALRCRTOCAADFFWELDPDWPSPAQASGAGALRDLSPFGGLLRRAACL 120  
Db 61 ALRSPAALRALRCRTOCAADFFWELDPDWPSPAQASGAGALRDLSPFGGLLRRAACL 120  
QY 121 RRCLGPPAAHSLSEMELEFRKSPYNYQVAYFKINKLEKAVAAATFFVGNPEHMEMQ 180

Sequence 2831, Ap  
Sequence 51334, A  
Sequence 64452, A  
Sequence 54435, A  
Sequence 79, Appl  
Sequence 6269, Ap  
Sequence 4501, Ap  
Sequence 143391,  
Sequence 2, Appli  
Sequence 64808, A  
Sequence 137691,  
Sequence 3528, Ap  
Sequence 8528, Ap  
Sequence 2363, Ap  
Sequence 9979, Ap  
Sequence 24, Appl  
Sequence 185927,  
Sequence 51318, A  
Sequence 544, App  
Sequence 32716, A  
Sequence 163956,  
Sequence 179700,  
Sequence 8355, Ap  
Sequence 11430, A  
Sequence 223454,  
Sequence 4597, Ap  
Sequence 133363,  
Sequence 8464, Ap  
Sequence 142593,  
Sequence 198322,

Db 121 RRLGPPAAHSLSEMELEFRKSPYNYLQVAFKINKLEKAVAAAHFTFFVGNPEHMEMQ 180  
 Qy 181 QNLDDYQTMGSKVKEADFKDLETPHMQEFLGVRLVLYSEEQPOQAVPHLEALQOEFVAYE 240  
 Db 181 QNLDDYQTMGSKVKEADFKDLETPHMQEFLGVRLVLYSEEQPOQAVPHLEALQOEFVAYE 240  
 Qy 241 ECRALCEGPDYDGYNYLQVAFKINKLEKAVAAAHFTFFVGNPEHMEMQ 300  
 Db 241 ECRALCEGPDYDGYNYLQVAFKINKLEKAVAAAHFTFFVGNPEHMEMQ 300  
 Qy 301 PSHYNYLQFAYNIGNYTQAGECAKTYLLFFPNDEVNQNLAAYAAMLGEEHTRSIGPRE 360  
 Db 301 PSHYNYLQFAYNIGNYTQAGECAKTYLLFFPNDEVNQNLAAYAAMLGEEHTRSIGPRE 360  
 Qy 361 SAKYRQRSLLKELLLFPAYDVGIPFVDPDSWTPPEEVI PKRLOEKOKSERETAVRISQE 420  
 Db 361 SAKYRQRSLLKELLLFPAYDVGIPFVDPDSWTPPEEVI PKRLOEKOKSERETAVRISQE 420  
 Qy 421 IGNLMKEIETLVEEKTESLDVSLTRREGGPLYEGISLTWNSKLLNGYQVVMWDGVID 480  
 Db 421 IGNLMKEIETLVEEKTESLDVSLTRREGGPLYEGISLTWNSKLLNGYQVVMWDGVID 480  
 Qy 481 HECQELQRLTNVAATSGDGYRGQTS PHTPNEKFGYGVTVFKALKLGQEGKVPLOSALHYN 540  
 Db 481 HECQELQRLTNVAATSGDGYRGQTS PHTPNEKFGYGVTVFKALKLGQEGKVPLOSALHYN 540  
 Qy 541 VTEKVRRLMESYFRDLTPLYFSYSHLVCRTAIEVQAEKDDSHPHVHVDNCILNAETLVC 600  
 Db 541 VTEKVRRLMESYFRDLTPLYFSYSHLVCRTAIEVQAEKDDSHPHVHVDNCILNAETLVC 600  
 Qy 601 VKEPPAYTFRDYSAILYLNGDFDGNFYFTLDAKTVAETVQPCGKRAVFGSSGTENPHG 660  
 Db 601 VKEPPAYTFRDYSAILYLNGDFDGNFYFTLDAKTVAETVQPCGKRAVFGSSGTENPHG 660  
 Qy 661 VKAVTRGQRCALWFTLDPHRSERDRVQADDLVKMLFSPPEMDLSQEQPLDAQGGPPEP 720  
 Db 661 VKAVTRGQRCALWFTLDPHRSERDRVQADDLVKMLFSPPEMDLSQEQPLDAQGGPPEP 720

RESULT 2  
 US-10-257-174-26  
 ; Sequence 26, Application US/10257174  
 ; Publication No. US20040034194A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Agarwal, Pankaj  
 ; APPLICANT: Murdoch, Paul R.  
 ; APPLICANT: Rizvi, Safia K.  
 ; APPLICANT: Smith, Randall F.  
 ; APPLICANT: Xiang, Zhaoying  
 ; TITLE OF INVENTION: NOVEL COMPOUNDS  
 ; FILE REFERENCE: GP50022  
 ; CURRENT APPLICATION NUMBER: US/10/257,174  
 ; CURRENT FILING DATE: 2002-10-10  
 ; PRIOR APPLICATION NUMBER: PCT/US01/11797  
 ; PRIOR FILING DATE: 2001-04-11  
 ; PRIOR APPLICATION NUMBER: 60/196,603  
 ; PRIOR FILING DATE: 2000-04-13  
 ; PRIOR APPLICATION NUMBER: 60/199,417  
 ; PRIOR FILING DATE: 2000-04-24  
 ; NUMBER OF SEQ ID NOS: 48  
 ; SOFTWARE: FastSeq for Windows Version 3.0  
 ; SEQ ID NO 26  
 ; LENGTH: 736  
 ; TYPE: PRT  
 ; ORGANISM: Homo sapiens  
 US-10-257-174-26  
 Query Match 58.8%; Score 433; DB 12; Length 736;  
 Best Local Similarity 99.6%; Pred. No. 0;

Matches 733; Conservative 0; Mismatches 3; Indels 0; Gaps 0;  
 Qy 1 MAVRALKLLTLLAVAAASQAEVSEAGMGWTPDLLFAEGTAAAYARGDWPGVVLWNER 60  
 Db 1 MAVRALKLLTLLAVAAASQAEVSEAGMGWTPDLLFAEGTAAAYARGDWPGVVLWNER 60  
 Qy 61 ALRSPAALRALRLRCRTQCAADFPWELDPDWSPPAQASGAGALRDLSPFFGGLLRRAACL 120  
 Db 61 ALRSPAALRALRLRCRTQCAADFPWELDPDWSPPAQASGAGALRDLSPFFGGLLRRAACL 120  
 Qy 121 RRLCGPPAAHSLSEMELEFRKSPYNYLQVAFKINKLEKAVAAAHFTFFVGNPEHMEMQ 180  
 Db 121 RRLCGPPAAHSLSEMELEFRKSPYNYLQVAFKINKLEKAVAAAHFTFFVGNPEHMEMQ 180  
 Qy 181 QNLDDYQTMGSKVKEADFKDLETPHMQEFLGVRLVLYSEEQPOQAVPHLEALQOEFVAYE 240  
 Db 181 QNLDDYQTMGSKVKEADFKDLETPHMQEFLGVRLVLYSEEQPOQAVPHLEALQOEFVAYE 240  
 Qy 241 ECRALCEGPDYDGYNYLQVAFKINKLEKAVAAAHFTFFVGNPEHMEMQ 300  
 Db 241 ECRALCEGPDYDGYNYLQVAFKINKLEKAVAAAHFTFFVGNPEHMEMQ 300  
 Qy 301 PSHYNYLQFAYNIGNYTQAGECAKTYLLFFPNDEVNQNLAAYAAMLGEEHTRSIGPRE 360  
 Db 301 PSHYNYLQFAYNIGNYTQAGECAKTYLLFFPNDEVNQNLAAYAAMLGEEHTRSIGPRE 360  
 Qy 361 SAKYRQRSLLKELLLFPAYDVGIPFVDPDSWTPPEEVI PKRLOEKOKSERETAVRISQE 420  
 Db 361 SAKYRQRSLLKELLLFPAYDVGIPFVDPDSWTPPEEVI PKRLOEKOKSERETAVRISQE 420  
 Qy 421 IGNLMKEIETLVEEKTESLDVSLTRREGGPLYEGISLTWNSKLLNGYQVVMWDGVID 480  
 Db 421 IGNLMKEIETLVEEKTESLDVSLTRREGGPLYEGISLTWNSKLLNGYQVVMWDGVID 480  
 Qy 481 HECQELQRLTNVAATSGDGYRGQTS PHTPNEKFGYGVTVFKALKLGQEGKVPLOSALHYN 540  
 Db 481 HECQELQRLTNVAATSGDGYRGQTS PHTPNEKFGYGVTVFKALKLGQEGKVPLOSALHYN 540  
 Qy 541 VTEKVRRLMESYFRDLTPLYFSYSHLVCRTAIEVQAEKDDSHPHVHVDNCILNAETLVC 600  
 Db 541 VTEKVRRLMESYFRDLTPLYFSYSHLVCRTAIEVQAEKDDSHPHVHVDNCILNAETLVC 600  
 Qy 601 VKEPPAYTFRDYSAILYLNGDFDGNFYFTLDAKTVAETVQPCGKRAVFGSSGTENPHG 660  
 Db 601 VKEPPAYTFRDYSAILYLNGDFDGNFYFTLDAKTVAETVQPCGKRAVFGSSGTENPHG 660  
 Qy 661 VKAVTRGQRCALWFTLDPHRSERDRVQADDLVKMLFSPPEMDLSQEQPLDAQGGPPEP 720  
 Db 661 VKAVTRGQRCALWFTLDPHRSERDRVQADDLVKMLFSPPEMDLSQEQPLDAQGGPPEP 720  
 Qy 721 AQESLSGSESKEPKDEL 736  
 Db 721 AQESLSGSESKEPKDEL 736

RESULT 3  
 US-10-312-352-33  
 ; Sequence 33, Application US/10312352  
 ; Publication No. US20040053824A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: INCYTE GENOMICS, INC.; TANG, Y. Tom  
 ; APPLICANT: YUE, Henry; AZIMZAI, Yalda  
 ; APPLICANT: HE, Ann; BATRA, Sajeev  
 ; APPLICANT: LO, Terence P.; NGUYEN, Dannie B.  
 ; APPLICANT: BURELL, John D.; MARCUS, Gregory A.  
 ; APPLICANT: ZINGLER, Kurt A.; GANDHI, Ameen R.  
 ; APPLICANT: LAL, Preeti G.; KEARNEY, Liam  
 ; APPLICANT: BURFORD, Neil; YAO, Monique G.  
 ; APPLICANT: CHAWLA, Narinder K.; ELLIOT, Vicki S.  
 ; APPLICANT: ARVIZU, Chandra S.; KHAN, Farrah A.  
 ; APPLICANT: BAUGHN, Mariah R.; HAFALIA, April, J.A.  
 ; APPLICANT: POLICKY, Jennifer L.; AU-YOUNG, Janice K.  
 ; APPLICANT: LU, Yan; BOROWSKY, Mark L.



APPLICANT: LU, Dyung Aina M.; RAMKUMAR, Javalaxmi  
APPLICANT: YANG, Juming; GURURAJAN, Rajagopal  
APPLICANT: WARREN, Bridget A.; GIETZEN, Kimberly J.  
APPLICANT: XU, Yuming; KALLICK, Deborah A.  
APPLICANT: LEE, Ernestine A.; THANGAVELU, Kavitha  
APPLICANT: DELEGANE, Angelo M.; LEE, Sally  
TITLE OF INVENTION: EXTRACELLULAR MATRIX AND CELL ADHESION MOLECULES  
FILE REFERENCE: PF-0794 USN  
CURRENT APPLICATION NUMBER: US/10/312,352  
CURRENT FILING DATE: 2002-12-18  
PRIOR APPLICATION NUMBER: PCT/US01/21067  
PRIOR FILING DATE: 2001-06-29  
PRIOR APPLICATION NUMBER: US 60/215,454  
PRIOR FILING DATE: 2000-06-30  
PRIOR APPLICATION NUMBER: US 60/219,462  
PRIOR FILING DATE: 2000-07-18  
PRIOR APPLICATION NUMBER: US 60/240,111  
PRIOR FILING DATE: 2000-10-12  
PRIOR APPLICATION NUMBER: US 60/240,106  
PRIOR FILING DATE: 2000-10-12  
PRIOR APPLICATION NUMBER: US 60/244,021  
PRIOR FILING DATE: 2000-10-27  
PRIOR APPLICATION NUMBER: US 60/248,887  
PRIOR FILING DATE: 2000-11-14  
PRIOR APPLICATION NUMBER: US 60/249,570  
PRIOR FILING DATE: 2000-11-16  
NUMBER OF SEQ ID NOS: 72  
SOFTWARE: PERL Program  
SEQ ID NO 33  
LENGTH: 736  
TYPE: PRT  
ORGANISM: Homo sapiens  
NAME/KEY: misc feature  
FEATURE:  
OTHER INFORMATION: Incyte ID No. US20040053824A1 6780147CD1  
US-10-312-352-33

Query Match 58.8%; Score 433; DB 12; Length 736;  
Best Local Similarity 99.6%; Pred. No. 0;  
Matches 733; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 MAVRALKLLTLLAVAAASQAQVESEAGWGWTPTDLLFAEGTAAAYARGDWPVGLSMER 60  
DB 1 MAVRALKLLTLLAVAAASQAQVESEAGWGWTPTDLLFAEGTAAAYARGDWPVGLSMER 60

QY 61 ALRSRAALRALRLRCRTQCAADFPWELDPDWSPPAASGAGALRDLSPFGGLLRAACL 120  
DB 61 ALRSRAALRALRLRCRTQCAADFPWELDPDWSPPAASGAGALRDLSPFGGLLRAACL 120

QY 121 RRCGLGPPAAHSLSEMELEFRKSPYNYLQVAYFKINKLEKAVAAAHTEFFVGNPEHMEMQ 180  
DB 121 RRCGLGPPAAHSLSEMELEFRKSPYNYLQVAYFKINKLEKAVAAAHTEFFVGNPEHMEMQ 180

QY 181 QNLDYQYTMGSKVEADFKDLETPHMQEFLGVRLYSEEQPQEAAPHLEAAALQYFVAYE 240  
DB 181 QNLDYQYTMGSKVEADFKDLETPHMQEFLGVRLYSEEQPQEAAPHLEAAALQYFVAYE 240

QY 241 ECRCALCEGPDYDGYNYLYNADLFOAITDHYIQLVNCQNCVTELASHPSREKPFDFL 300  
DB 241 ECRCALCEGPDYDGYNYLYNADLFOAITDHYIQLVNCQNCVTELASHPSREKPFDFL 300

QY 301 PSHYNYLOPYYNIGNYTOAGECAKTYLLFFPNDEVNQNLAAYAAAMLGEEHTRSGPRE 360  
DB 301 PSHYNYLOPYYNIGNYTOAGECAKTYLLFFPNDEVNQNLAAYAAAMLGEEHTRSGPRE 360

QY 361 SAKEYQRSLLLEKELLFFAYDVFGIPFVDPDSMTPEVTPKRLQEKQKSERETAVRISQE 420  
DB 361 SAKEYQRSLLLEKELLFFAYDVFGIPFVDPDSMTPEVTPKRLQEKQKSERETAVRISQE 420

QY 421 IGNLMKEITLVEEKTKESLDVSRLEGGPILLYEGISLTMSKLLNGYQRYVMGVISD 480  
DB 421 IGNLMKEITLVEEKTKESLDVSRLEGGPILLYEGISLTMSKLLNGYQRYVMGVISD 480

QY 481 HECQELQRLTNVAATSGDGYRGQTSPHTNEKFYGVTVFKALKLGOEGKVPLOSAHLYN 540  
DB 481 HECQELQRLTNVAATSGDGYRGQTSPHTNEKFYGVTVFKALKLGOEGKVPLOSAHLYN 540

QY 541 VTEKVRIMESYFRDLTPLYFSYSHLVCRITAEEVQAERKODSHPHVDNCILNAETLVC 600  
DB 541 VTEKVRIMESYFRDLTPLYFSYSHLVCRITAEEVQAERKODSHPHVDNCILNAETLVC 600

QY 601 VKEPPAYTFRDYSAIYLINGDPDGNFYFTELDAKTVTAEVQPCGRAVSGSGTENPHG 660  
DB 601 VKEPPAYTFRDYSAIYLINGDPDGNFYFTELDAKTVTAEVQPCGRAVSGSGTENPHG 660

QY 661 VKAVTRGORCAIALMFTLDPHRSERDRVQADDLVKMLFSPPEMDLSQSQPLDAQQPPPEP 720  
DB 661 VKAVTRGORCAIALMFTLDPHRSERDRVQADDLVKMLFSPPEMDLSQSQPLDAQQPPPEP 720

QY 721 AQESLSGSESKPKDEL 736  
DB 721 AQESLSGSESKPKDEL 736

RESULT 4  
US-10-045-815-2  
; Sequence 2, Application US/10045815  
; Publication No. US20020160498A1  
; GENERAL INFORMATION:  
; APPLICANT: Wadhwa, Renu  
; APPLICANT: Sugihara, Takashi  
; APPLICANT: Ohide, Akiko  
; TITLE OF INVENTION: TUMOR SUPPRESSOR GENE  
; FILE REFERENCE: 06501-091001  
; CURRENT APPLICATION NUMBER: US/10/045,815  
; CURRENT FILING DATE: 2001-10-26  
; PRIOR APPLICATION NUMBER: PCT/JP00/02731  
; PRIOR FILING DATE: 2000-04-26  
; PRIOR APPLICATION NUMBER: JP 11/118806  
; PRIOR FILING DATE: 1999-04-26  
; NUMBER OF SEQ ID NOS: 15  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 2  
; LENGTH: 363  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
US-10-045-815-2

Query Match 48.9%; Score 360; DB 13; Length 363;  
Best Local Similarity 100.0%; Pred. No. 0;  
Matches 360; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MAVRALKLLTLLAVAAASQAQVESEAGWGWTPTDLLFAEGTAAAYARGDWPVGLSMER 60  
DB 1 MAVRALKLLTLLAVAAASQAQVESEAGWGWTPTDLLFAEGTAAAYARGDWPVGLSMER 60

QY 61 ALRSRAALRALRLRCRTQCAADFPWELDPDWSPPAASGAGALRDLSPFGGLLRAACL 120  
DB 61 ALRSRAALRALRLRCRTQCAADFPWELDPDWSPPAASGAGALRDLSPFGGLLRAACL 120

QY 121 RRCGLGPPAAHSLSEMELEFRKSPYNYLQVAYFKINKLEKAVAAAHTEFFVGNPEHMEMQ 180  
DB 121 RRCGLGPPAAHSLSEMELEFRKSPYNYLQVAYFKINKLEKAVAAAHTEFFVGNPEHMEMQ 180

QY 181 QNLDYQYTMGSKVEADFKDLETPHMQEFLGVRLYSEEQPQEAAPHLEAAALQYFVAYE 240  
DB 181 QNLDYQYTMGSKVEADFKDLETPHMQEFLGVRLYSEEQPQEAAPHLEAAALQYFVAYE 240

QY 241 ECRCALCEGPDYDGYNYLYNADLFOAITDHYIQLVNCQNCVTELASHPSREKPFDFL 300  
DB 241 ECRCALCEGPDYDGYNYLYNADLFOAITDHYIQLVNCQNCVTELASHPSREKPFDFL 300

QY 301 PSHYNYLOPYYNIGNYTOAGECAKTYLLFFPNDEVNQNLAAYAAAMLGEEHTRSGPRE 360  
DB 301 PSHYNYLOPYYNIGNYTOAGECAKTYLLFFPNDEVNQNLAAYAAAMLGEEHTRSGPRE 360

```

RESULT 5
US-10-257-174-25
; Sequence 25, Application US/10257174
; Publication No. US2004003419A1
; GENERAL INFORMATION:
; APPLICANT: Agarwal, Pankaj
; APPLICANT: Murdock, Paul R.
; APPLICANT: Rizvi, Safia K.
; APPLICANT: Smith, Randall F.
; APPLICANT: Xiang, Zhaoxing
; TITLE OF INVENTION: NOVEL COMPOUNDS
; FILE REFERENCE: GP50022
; CURRENT APPLICATION NUMBER: US/10/257,174
; CURRENT FILING DATE: 2002-10-10
; PRIOR APPLICATION NUMBER: PCT/US01/11797
; PRIOR FILING DATE: 2001-04-11
; PRIOR APPLICATION NUMBER: 60/196,603
; PRIOR FILING DATE: 2000-04-13
; PRIOR APPLICATION NUMBER: 60/199,417
; PRIOR FILING DATE: 2000-04-24
; NUMBER OF SEQ ID NOS: 48
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 25
; LENGTH: 708
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-257-174-25

Query Match      37.4%; Score 275; DB 12; Length 708;
Best Local Similarity 99.7%; Pred. No. 5.6e-254;
Matches 375; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 361 SAKYRORSILLEKELLIFAYDVFGIPVDPSWTPPEVVKRLOEKOKSERETAVRISQ 420
DB 333 SAKYRORSILLEKELLIFAYDVFGIPVDPSWTPPEVVKRLOEKOKSERETAVRISQ 392

QY 421 IGMLMKIETLVEKTKESLDVSRLTREGGPLYEGISLTWNSKLLNGYORVMDGVISD 480
DB 393 IGMLMKIETLVEKTKESLDVSRLTREGGPLYEGISLTWNSKLLNGSRVMDGVISD 452

QY 481 HECQELQRLTNVAATSGDGRGTSPHTPNEKFGVTVFVXALXGQEGKVPLOSAHLYN 540
DB 453 HECQELQRLTNVAATSGDGRGTSPHTPNEKFGVTVFVXALXGQEGKVPLOSAHLYN 512

QY 541 VTKVRIMESYFLDTPLYFSYSHLVCRTAIEVQAEKDDSHPVHVDNCILNAETLVC 600
DB 513 VTKVRIMESYFLDTPLYFSYSHLVCRTAIEVQAEKDDSHPVHVDNCILNAETLVC 572

QY 601 VKEPPAYTRDYSAILYNGDFDGNFYFTELDAAKTVTAEVQPCGRAGVFGSGTENPHG 660
DB 573 VKEPPAYTRDYSAILYNGDFDGNFYFTELDAAKTVTAEVQPCGRAGVFGSGTENPHG 632

QY 661 VKAVTRGRCALWFTLDPHRSERDRVQADDLVKMLFSPPEMDLSQEQPLDAQCGPPEP 720
DB 633 VKAVTRGRCALWFTLDPHRSERDRVQADDLVKMLFSPPEMDLSQEQPLDAQCGPPEP 692

QY 721 AQESLSGSESXPKDEL 736
DB 693 AQESLSGSESXPKDEL 708

```

```

RESULT 6
US-09-833-245-1251
; Sequence 1251, Application US/09833245
; Publication No. US20040010134A1
; GENERAL INFORMATION:
; APPLICANT: Human Genome Sciences, Inc.
; TITLE OF INVENTION: Albumin Fusion Proteins
; FILE REFERENCE: PF546PCT
; CURRENT APPLICATION NUMBER: US/09/833,245
; CURRENT FILING DATE: 2001-04-12
; PRIOR APPLICATION NUMBER: 60/229,358
; PRIOR FILING DATE: 2000-04-12
; PRIOR APPLICATION NUMBER: 60/256,931
; PRIOR FILING DATE: 2000-12-21
; PRIOR APPLICATION NUMBER: 60/199,384
; PRIOR FILING DATE: 2000-04-25
; NUMBER OF SEQ ID NOS: 2267
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 1250
; LENGTH: 173
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-833-245-1250

Query Match      13.7%; Score 101; DB 11; Length 173;
Best Local Similarity 100.0%; Pred. No. 5.9e-88;
Matches 101; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MAVRAKLLTLLAVVAAASQAEVSEAGWGMVTPDLLFAEGTAAAYARGDWPVGVLSMER 60
DB 1 MAVRAKLLTLLAVVAAASQAEVSEAGWGMVTPDLLFAEGTAAAYARGDWPVGVLSMER 60

```

```

; PRIOR FILING DATE: 2000-04-12
; PRIOR APPLICATION NUMBER: 60/256,931
; PRIOR FILING DATE: 2000-12-21
; PRIOR APPLICATION NUMBER: 60/199,384
; PRIOR FILING DATE: 2000-04-25
; NUMBER OF SEQ ID NOS: 2267
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 1251
; LENGTH: 359
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-833-245-1251

Query Match      29.8%; Score 219; DB 11; Length 359;
Best Local Similarity 99.7%; Pred. No. 1.2e-200;
Matches 319; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 MAVRAKLLTLLAVVAAASQAEVSEAGWGMVTPDLLFAEGTAAAYARGDWPVGVLSMER 60
DB 1 MAVRAKLLTLLAVVAAASQAEVSEAGWGMVTPDLLFAEGTAAAYARGDWPVGVLSMER 60

QY 61 ALRSPAALRALRLCRTOCAADFWEKLDPDWSPQAASGAGALRDLSPFGGLLRRACL 120
DB 61 ALRSPAALRALRLCRTOCAADFWEKLDPDWSPQAASGAGALRDLSPFGGLLRRACL 120

QY 121 RRLGPPAAHSLSEEMELEFRKRSPPYNYQVAYFKINKLEKAVAAAHFTFFVGNPEHMEMQ 180
DB 121 RRLGPPAAHSLSEEMELEFRKRSPPYNYQVAYFKINKLEKAVAAAHFTFFVGNPEHMEMQ 180

QY 181 QNLDDYQTMGVKAEADFKDLETOHPMQEFLGVLYSEEQPQEAHPHLEALQEVFVAYE 240
DB 181 QNLDDYQTMGVKAEADFKDLETOHPMQEFLGVLYSEEQPQEAHPHLEALQEVFVAYE 240

QY 241 ECRALECEPYDYDGVNYLEYNADLFQAITDHYIQLVNLCKQNCVTELASHPSREKPFEDFL 300
DB 241 ECRALECEPYDYDGVNYLEYNADLFQAITDHYIQLVNLCKQNCVTELASHPSREKPFEDFL 300

QY 301 PSHYNYLOFAYNYIGNYTQA 320
DB 301 PSHYNYLOFAYNYIGNYTQA 320

```

```

RESULT 7
US-09-833-245-1250
; Sequence 1250, Application US/09833245
; Publication No. US20040010134A1
; GENERAL INFORMATION:
; APPLICANT: Human Genome Sciences, Inc.
; TITLE OF INVENTION: Albumin Fusion Proteins
; FILE REFERENCE: PF546PCT
; CURRENT APPLICATION NUMBER: US/09/833,245
; CURRENT FILING DATE: 2001-04-12
; PRIOR APPLICATION NUMBER: 60/229,358
; PRIOR FILING DATE: 2000-04-12
; PRIOR APPLICATION NUMBER: 60/256,931
; PRIOR FILING DATE: 2000-12-21
; PRIOR APPLICATION NUMBER: 60/199,384
; PRIOR FILING DATE: 2000-04-25
; NUMBER OF SEQ ID NOS: 2267
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 1250
; LENGTH: 173
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-833-245-1250

Query Match      13.7%; Score 101; DB 11; Length 173;
Best Local Similarity 100.0%; Pred. No. 5.9e-88;
Matches 101; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MAVRAKLLTLLAVVAAASQAEVSEAGWGMVTPDLLFAEGTAAAYARGDWPVGVLSMER 60
DB 1 MAVRAKLLTLLAVVAAASQAEVSEAGWGMVTPDLLFAEGTAAAYARGDWPVGVLSMER 60

```

Qy	61	ALRGAALRALRLRCRTQCAADFPPWELDPWSPPAQASGA	101
Dd	61	ALRGAALRALRLRCRTQCAADFPPWELDPWSPPAQASGA	101

; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies  
 ; FILE REFERENCE: P0202  
 ; CURRENT APPLICATION NUMBER: US/09/764,875  
 ; CURRENT FILING DATE: 2001-01-17  
 ; Prior application data removed - consult PALM or file wrapper  
 ; NUMBER OF SEQ ID NOS: 1249  
 ; SOFTWARE: PatentIn Ver. 2.0  
 ; SEQ ID NO 1106  
 ; LENGTH: 267  
 ; TYPE: PRT  
 ; ORGANISM: Homo sapiens  
 ; FEATURE:  
 ; NAME/KEY: SITE  
 ; LOCATION: (6)  
 ; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids  
 US-09-764-875-1106

Query Match 1.8%; Score 13; DB 11; Length 267;  
 Best Local Similarity 100.0%; Pred. No. 0.0015;  
 Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 602 KEPPAYTRDYSA 614  
 Db 154 KEPPAYTRDYSA 166

RESULT 13

US-10-012-697-1493  
 ; Sequence 1493, Application US/10012697  
 ; Publication No. US20030215803A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Escobedo, Jaime  
 ; APPLICANT: Garcia, Pablo Dominguez  
 ; APPLICANT: Kassam, Altaf  
 ; APPLICANT: Lamson, George  
 ; APPLICANT: Scott, Beth  
 ; APPLICANT: Drmanac, Radoje  
 ; APPLICANT: Crkvenjakov, Radomir  
 ; APPLICANT: Dickson, Mark  
 ; APPLICANT: Drmanac, Snezana  
 ; APPLICANT: Labat, Ivan  
 ; APPLICANT: Leshkowitz, Dena  
 ; APPLICANT: Kita, David  
 ; APPLICANT: Garcia, Veronica  
 ; APPLICANT: Jones, Lee William  
 ; APPLICANT: Stache-Crain, Birgit  
 ; TITLE OF INVENTION: HUMAN GENES AND GENE EXPRESSION PRODUCTS  
 ; FILE REFERENCE: 2300-16252  
 ; CURRENT APPLICATION NUMBER: US/10/012,697  
 ; CURRENT FILING DATE: 2003-01-21  
 ; PRIOR APPLICATION NUMBER: 60/254,648  
 ; PRIOR FILING DATE: 2000-12-07  
 ; PRIOR APPLICATION NUMBER: 60/275,668  
 ; PRIOR FILING DATE: 2001-03-13  
 ; NUMBER OF SEQ ID NOS: 1568  
 ; SOFTWARE: FastSeq for Windows Version 4.0  
 ; SEQ ID NO 1493  
 ; LENGTH: 282  
 ; TYPE: PRT  
 ; ORGANISM: Homo sapiens  
 US-10-012-697-1493

Query Match 1.8%; Score 13; DB 15; Length 282;  
 Best Local Similarity 100.0%; Pred. No. 0.0016;  
 Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 602 KEPPAYTRDYSA 614  
 Db 169 KEPPAYTRDYSA 181

RESULT 14

US-10-094-749-2560  
 ; Sequence 2560, Application US/10094749  
 ; Publication No. US20030219741A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: ISOGAI, TAKAO  
 ; APPLICANT: SUGIYAMA, TOMOYASU  
 ; APPLICANT: OTSUKI, TETSUJI  
 ; APPLICANT: WAKAMATSU, AI  
 ; APPLICANT: SATO, HIROYUKI  
 ; APPLICANT: ISHII, SHIZUKO  
 ; APPLICANT: YAMAMOTO, JUN-ICHI  
 ; APPLICANT: ISONO, YUUKO  
 ; APPLICANT: HIO, YURI  
 ; APPLICANT: OTSUKA, KAORU  
 ; APPLICANT: NAGAI, KEIICHI  
 ; APPLICANT: IRIE, RYOTARO  
 ; APPLICANT: TAMECHIKA, ICHIRO  
 ; APPLICANT: SEKI, NAOHICO  
 ; APPLICANT: YOSHIKAWA, TSUTOMU  
 ; APPLICANT: OTSUKA, MOTOKYUKI  
 ; APPLICANT: NAGAHARI, KENJI  
 ; APPLICANT: MASUHO, YASUHIKO  
 ; TITLE OF INVENTION: NOVEL FULL-LENGTH cDNA  
 ; FILE REFERENCE: 084335/0160  
 ; CURRENT APPLICATION NUMBER: US/10/094,749  
 ; CURRENT FILING DATE: 2002-03-12  
 ; PRIOR APPLICATION NUMBER: 60/350,435  
 ; PRIOR FILING DATE: 2002-01-24  
 ; PRIOR APPLICATION NUMBER: JP 2001-328381  
 ; PRIOR FILING DATE: 2001-09-14  
 ; NUMBER OF SEQ ID NOS: 3381  
 ; SOFTWARE: PatentIn Ver. 2.1  
 ; SEQ ID NO 2560  
 ; LENGTH: 708  
 ; TYPE: PRT  
 ; ORGANISM: Homo sapiens  
 US-10-094-749-2560

Query Match 1.8%; Score 13; DB 15; Length 708;  
 Best Local Similarity 100.0%; Pred. No. 0.0036;  
 Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 602 KEPPAYTRDYSA 614  
 Db 595 KEPPAYTRDYSA 607

RESULT 15

US-09-572-404B-1740  
 ; Sequence 1740, Application US/09572404B  
 ; Publication No. US20030078374A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Proteom Ltd  
 ; TITLE OF INVENTION: Complementary peptide ligands from the human genome  
 ; FILE REFERENCE: Human patent  
 ; CURRENT APPLICATION NUMBER: US/09/572,404B  
 ; CURRENT FILING DATE: 2000-05-17  
 ; NUMBER OF SEQ ID NOS: 4203  
 ; SOFTWARE: ProtPatent version 1.0  
 ; SEQ ID NO 1740  
 ; LENGTH: 10  
 ; TYPE: PRT  
 ; ORGANISM: Homo Sapiens  
 ; FEATURE:  
 ; OTHER INFORMATION: sequence located in NOP56 at 110-119 and may interact with Sequenc  
 ; OTHER INFORMATION: in this patent.  
 US-09-572-404B-1740

Query Match 1.1%; Score 8; DB 10; Length 10;  
 Best Local Similarity 100.0%; Pred. No. 4.9;  
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 116 RAACLRRRC 123

Db           |||||||  
              3 RAAOLREC 10

Search completed: July 18, 2004, 10:02:04  
Job time : 54 secs

This Page Blank (uspto)

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: July 18, 2004, 09:41:18 ; Search time 685 Seconds  
(without alignments)  
335.832 Million cell updates/sec

Title: US-10-045-815-4

Perfect score: 3870  
Sequence: 1 MAVRAKLKLTLLAVVAAS.....PPEPAQESLSGSESKPKDEL 736

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 1285345 seqs, 312560633 residues

Total number of hits satisfying chosen parameters: 1285345

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : Published Applications AA:\*

- 1: /cgn2\_6/prodata/2/pubpaa/US07\_PUBCOMB.pep.\*
- 2: /cgn2\_6/prodata/2/pubpaa/PCT\_NEW\_PUB.pep.\*
- 3: /cgn2\_6/prodata/2/pubpaa/US06\_NEW\_PUB.pep.\*
- 4: /cgn2\_6/prodata/2/pubpaa/US06\_PUBCOMB.pep.\*
- 5: /cgn2\_6/prodata/2/pubpaa/US07\_NEW\_PUB.pep.\*
- 6: /cgn2\_6/prodata/2/pubpaa/PCTUS\_PUBCOMB.pep.\*
- 7: /cgn2\_6/prodata/2/pubpaa/US08\_NEW\_PUB.pep.\*
- 8: /cgn2\_6/prodata/2/pubpaa/US08\_PUBCOMB.pep.\*
- 9: /cgn2\_6/prodata/2/pubpaa/US09A\_PUBCOMB.pep.\*
- 10: /cgn2\_6/prodata/2/pubpaa/US09B\_PUBCOMB.pep.\*
- 11: /cgn2\_6/prodata/2/pubpaa/US09C\_PUBCOMB.pep.\*
- 12: /cgn2\_6/prodata/2/pubpaa/US09\_NEW\_PUB.pep.\*
- 13: /cgn2\_6/prodata/2/pubpaa/US10A\_PUBCOMB.pep.\*
- 14: /cgn2\_6/prodata/2/pubpaa/US10B\_PUBCOMB.pep.\*
- 15: /cgn2\_6/prodata/2/pubpaa/US10C\_PUBCOMB.pep.\*
- 16: /cgn2\_6/prodata/2/pubpaa/US10\_NEW\_PUB.pep.\*
- 17: /cgn2\_6/prodata/2/pubpaa/US60\_NEW\_PUB.pep.\*
- 18: /cgn2\_6/prodata/2/pubpaa/US60\_PUBCOMB.pep.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	3870	100.0	736	13	US-10-045-815-4
2	3846	99.4	736	12	Sequence 26, Appl
3	3846	99.4	736	12	Sequence 33, Appl
4	3572	92.3	708	13	US-10-312-352-33
5	3316	85.7	747	13	US-10-257-174-25
6	2384	61.6	542	13	US-10-045-815-6
7	1911	49.4	363	13	US-10-045-815-8
8	1800	46.5	359	11	US-09-833-245-1251
9	1751	45.2	708	15	US-10-094-749-2560
10	1473	38.1	736	15	US-10-094-886-24
11	906.5	23.4	282	15	US-10-012-697-1493
12	836.5	21.6	267	9	US-09-764-868-736
13	836.5	21.6	267	11	US-09-764-875-1106
14	657	17.0	173	11	US-09-833-245-1250
15	612	15.8	219	11	US-09-764-875-815

16	567	14.7	401	9	US-09-729-674-2	Sequence 2, Appli
17	501	12.9	528	15	US-10-094-749-2363	Sequence 2363, Ap
18	383.5	9.9	153	12	US-10-276-774-1567	Sequence 1567, Ap
19	274	7.1	262	15	US-10-108-260A-4501	Sequence 4501, Ap
20	183.5	4.7	104	15	US-10-108-260A-3928	Sequence 3928, Ap
21	177	4.6	241	12	US-10-424-599-224808	Sequence 224808,
22	148.5	3.8	603	15	US-10-289-762-1011	Sequence 1011, Ap
23	135	3.5	1207	12	US-10-250-823-8	Sequence 8, Appli
24	135	3.5	1207	12	US-10-250-823-10	Sequence 10, Appl
25	129	3.3	227	9	US-09-782-980-79	Sequence 79, Appl
26	126.5	3.3	1154	12	US-10-250-823-4	Sequence 4, Appli
27	126.5	3.3	1154	12	US-10-250-823-6	Sequence 6, Appli
28	123	3.2	2771	16	US-10-437-963-195406	Sequence 195406,
29	122	3.2	1529	16	US-10-389-566-2103	Sequence 2103, Ap
30	119.5	3.1	1905	15	US-10-259-194A-86	Sequence 86, Appl
31	118.5	3.1	847	12	US-10-262-511-8	Sequence 8, Appli
32	118.5	3.1	2053	13	US-10-017-216-2	Sequence 2, Appli
33	118.5	3.1	2053	14	US-10-325-430-12	Sequence 12, Appl
34	118	3.0	1441	15	US-10-412-897-3	Sequence 3, Appli
35	118	3.0	1597	12	US-09-564-956-41	Sequence 41, Appl
36	118	3.0	1597	13	US-10-017-216-6	Sequence 6, Appli
37	118	3.0	1641	12	US-09-964-956-40	Sequence 40, Appl
38	118	3.0	1641	13	US-10-017-216-5	Sequence 5, Appli
39	118	3.0	2055	13	US-10-017-216-4	Sequence 4, Appli
40	117	3.0	832	12	US-10-262-511-6	Sequence 6, Appli
41	117	3.0	1286	12	US-09-964-956-38	Sequence 38, Appl
42	117	3.0	1286	13	US-10-017-216-7	Sequence 7, Appli
43	117	3.0	1958	13	US-10-028-946-4	Sequence 4, Appli
44	117	3.0	2053	12	US-09-964-956-11	Sequence 11, Appl
45	117	3.0	2053	12	US-10-262-511-2	Sequence 2, Appli

## ALIGNMENTS

RESULT 1  
US-10-045-815-4  
; Sequence 4, Application US/10045815  
; Publication No. US20020160498A1  
; GENERAL INFORMATION:  
; APPLICANT: Wadwa, Renu  
; APPLICANT: Sugihara, Takashi  
; APPLICANT: Ohide, Akiko  
; TITLE OF INVENTION: TUMOR SUPPRESSOR GENE  
; FILE REFERENCE: 06501-091001  
; CURRENT APPLICATION NUMBER: US/10/045,815  
; CURRENT FILING DATE: 2001-10-26  
; PRIOR APPLICATION NUMBER: PCT/JF00/02731  
; PRIOR FILING DATE: 2000-04-26  
; PRIOR APPLICATION NUMBER: JP 11/118806  
; PRIOR FILING DATE: 1999-04-26  
; NUMBER OF SEQ ID NOS: 15  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 4  
; LENGTH: 736  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
US-10-045-815-4

Query Match	100.0%	Score	3870;	DB	13;	Length	736;
Best Local Similarity	100.0%;	Pred. No.	0;				
Matches	736;	Conservative	0;	Mismatches	0;	Indels	0;
Gaps	0;						
QY	1	MAVRALKLTLLAVVAASQA	VESEAGWGVTPDLLFAEGTAAYARGDPGVVLSMER	60			
Db	1	MAVRALKLTLLAVVAASQA	VESEAGWGVTPDLLFAEGTAAYARGDPGVVLSMER	60			
QY	61	ALRSPAALRALRLRCRTCAAD	FPWELDPWSPQAASGAGALRDLSPFGGLLRAACL	120			
Db	61	ALRSPAALRALRLRCRTCAAD	FPWELDPWSPQAASGAGALRDLSPFGGLLRAACL	120			
QY	121	RRCLQPPAAHSLSMELEFRKSP	PNYLVAYFKINKLEKAVAAAHITFFVGNPEHMEMQ	180			
Db	121	RRCLQPPAAHSLSMELEFRKSP	PNYLVAYFKINKLEKAVAAAHITFFVGNPEHMEMQ	180			

Db 121 RRCIGPPAAHSLSEMELEFRKRSPPNYLVQVAFKINKLEKAVAAAHTFFVGNPEHMEMQ 180  
Qy 181 QNLDDYQTMGSKVKEADFKDLETPHMQEFLGLVRLYSEEPQAVPHLEAALQEFVAYE 240  
Db 181 QNLDDYQTMGSKVKEADFKDLETPHMQEFLGLVRLYSEEPQAVPHLEAALQEFVAYE 240  
Qy 241 ECRALCEGPDYDGYNYLEYNADLFOAITDHYIQVLNCKQNCVTELASHPSREKPFEDFL 300  
Db 241 ECRALCEGPDYDGYNYLEYNADLFOAITDHYIQVLNCKQNCVTELASHPSREKPFEDFL 300  
Qy 301 PSHYNYLQFAYYNTGNTOGECACKTYLLFPFNDENVNQNLAAYAAAMLGEHTRSIGPRE 360  
Db 301 PSHYNYLQFAYYNTGNTOGECACKTYLLFPFNDENVNQNLAAYAAAMLGEHTRSIGPRE 360  
Qy 361 SAKYRQSRSLLEKELLFFAYDVFGIPFVDPDSWTPPEEVIKRLQEKQKSERETAVRISQE 420  
Db 361 SAKYRQSRSLLEKELLFFAYDVFGIPFVDPDSWTPPEEVIKRLQEKQKSERETAVRISQE 420  
Qy 421 IGNLKMEIETLVEKTKESLDVSRLTREGGPLLVEGSLTMNSKLLNGYORVMDGVISD 480  
Db 421 IGNLKMEIETLVEKTKESLDVSRLTREGGPLLVEGSLTMNSKLLNGYORVMDGVISD 480  
Qy 481 HECQELQRLTNVAATSGDGYRGQTSPTPNEKFGYVTVFKALKLQGBGKVPLOSAHLYN 540  
Db 481 HECQELQRLTNVAATSGDGYRGQTSPTPNEKFGYVTVFKALKLQGBGKVPLOSAHLYN 540  
Qy 541 VTEKVRIMESYFRDLTPLYFSYSHLVCRTAIEVQAEKDDSHPHVHVDNCILNAETLVC 600  
Db 541 VTEKVRIMESYFRDLTPLYFSYSHLVCRTAIEVQAEKDDSHPHVHVDNCILNAETLVC 600  
Qy 601 VKEPPATFRDYSAILYLNGDFDGNFYFTELDKTTVAEVOQCGRAVGFSSGTENPHG 660  
Db 601 VKEPPATFRDYSAILYLNGDFDGNFYFTELDKTTVAEVOQCGRAVGFSSGTENPHG 660  
Qy 661 VKAVTRGORCAIALWFTLDPHRSERDRVQADDLVKMLFSPPEMDLSQEQPLDAQOGPPEP 720  
Db 661 VKAVTRGORCAIALWFTLDPHRSERDRVQADDLVKMLFSPPEMDLSQEQPLDAQOGPPEP 720  
Qy 721 AQESLSGSESXPKDEL 736  
Db 721 AQESLSGSESXPKDEL 736

## RESULT 2

US-10-257-174-26  
; Sequence 26, Application US/10257174  
; Publication No. US20040034194A1  
; GENERAL INFORMATION:  
; APPLICANT: Agarwal, Pankaj  
; APPLICANT: Murdock, Paul R.  
; APPLICANT: Rizvi, Safia K.  
; APPLICANT: Smith, Randall F.  
; APPLICANT: Xiang, Zhaoying  
; TITLE OF INVENTION: NOVEL COMPOUNDS  
; FILE REFERENCE: GP50022  
; CURRENT APPLICATION NUMBER: US/10/257,174  
; CURRENT FILING DATE: 2002-10-10  
; PRIOR APPLICATION NUMBER: PCT/US01/11797  
; PRIOR FILING DATE: 2001-04-11  
; PRIOR APPLICATION NUMBER: 60/196,603  
; PRIOR FILING DATE: 2000-04-13  
; PRIOR APPLICATION NUMBER: 60/199,417  
; PRIOR FILING DATE: 2000-04-24  
; NUMBER OF SEQ ID NOS: 48  
; SOFTWARE: FastSeq for Windows Version 3.0  
; SEQ ID NO 26  
; LENGTH: 736  
; TYPE: PRT  
; ORGANISM: Homo sapiens

US-10-257-174-26

Query Match

Best Local Similarity

99.4%; Score 3846; DB 12; Length 736;

99.6%; Pred. No. 0;

Matches 733; Conservative 0; Mismatches 3; Indels 0; Gaps 0;  
Qy 1 MAVRALKLLTLLAVAAAASQAEVESEAGMGWTPDLLFAEGTAAYARGDMPGVVLSMER 60  
Db 1 MAVRALKLLTLLAVAAAASQAEVESEAGMGWTPDLLFAEGTAAYARGDMPGVVLSMER 60  
Qy 61 ALRSPAAALRALRLCRTOCAADFFWELDPDWSPPAQASGAGALDELDFGGLLRRACL 120  
Db 61 ALRSPAAALRALRLCRTOCAADFFWELDPDWSPPAQASGAGALDELDFGGLLRRACL 120  
Qy 121 RRCIGPPAAHSLSEMELEFRKRSPPNYLVQVAFKINKLEKAVAAAHTFFVGNPEHMEMQ 180  
Db 121 RRCIGPPAAHSLSEMELEFRKRSPPNYLVQVAFKINKLEKAVAAAHTFFVGNPEHMEMQ 180  
Qy 181 QNLDDYQTMGSKVKEADFKDLETPHMQEFLGLVRLYSEEPQAVPHLEAALQEFVAYE 240  
Db 181 QNLDDYQTMGSKVKEADFKDLETPHMQEFLGLVRLYSEEPQAVPHLEAALQEFVAYE 240  
Qy 241 ECRALCEGPDYDGYNYLEYNADLFOAITDHYIQVLNCKQNCVTELASHPSREKPFEDFL 300  
Db 241 ECRALCEGPDYDGYNYLEYNADLFOAITDHYIQVLNCKQNCVTELASHPSREKPFEDFL 300  
Qy 301 PSHYNYLQFAYYNTGNTOGECACKTYLLFPFNDENVNQNLAAYAAAMLGEHTRSIGPRE 360  
Db 301 PSHYNYLQFAYYNTGNTOGECACKTYLLFPFNDENVNQNLAAYAAAMLGEHTRSIGPRE 360  
Qy 361 SAKYRQSRSLLEKELLFFAYDVFGIPFVDPDSWTPPEEVIKRLQEKQKSERETAVRISQE 420  
Db 361 SAKYRQSRSLLEKELLFFAYDVFGIPFVDPDSWTPPEEVIKRLQEKQKSERETAVRISQE 420  
Qy 421 IGNLKMEIETLVEKTKESLDVSRLTREGGPLLVEGSLTMNSKLLNGYORVMDGVISD 480  
Db 421 IGNLKMEIETLVEKTKESLDVSRLTREGGPLLVEGSLTMNSKLLNGYORVMDGVISD 480  
Qy 481 HECQELQRLTNVAATSGDGYRGQTSPTPNEKFGYVTVFKALKLQGBGKVPLOSAHLYN 540  
Db 481 HECQELQRLTNVAATSGDGYRGQTSPTPNEKFGYVTVFKALKLQGBGKVPLOSAHLYN 540  
Qy 541 VTEKVRIMESYFRDLTPLYFSYSHLVCRTAIEVQAEKDDSHPHVHVDNCILNAETLVC 600  
Db 541 VTEKVRIMESYFRDLTPLYFSYSHLVCRTAIEVQAEKDDSHPHVHVDNCILNAETLVC 600  
Qy 601 VKEPPATFRDYSAILYLNGDFDGNFYFTELDKTTVAEVOQCGRAVGFSSGTENPHG 660  
Db 601 VKEPPATFRDYSAILYLNGDFDGNFYFTELDKTTVAEVOQCGRAVGFSSGTENPHG 660  
Qy 661 VKAVTRGORCAIALWFTLDPHRSERDRVQADDLVKMLFSPPEMDLSQEQPLDAQOGPPEP 720  
Db 661 VKAVTRGORCAIALWFTLDPHRSERDRVQADDLVKMLFSPPEMDLSQEQPLDAQOGPPEP 720  
Qy 721 AQESLSGSESXPKDEL 736  
Db 721 AQESLSGSESXPKDEL 736

## RESULT 3

US-10-312-352-33  
; Sequence 33, Application US/10312352  
; Publication No. US20040053824A1  
; GENERAL INFORMATION:  
; APPLICANT: INCYTE GENOMICS, INC.; TANG, Y. Tom  
; APPLICANT: YUE, Henry; AZIMZAI, Yalda  
; APPLICANT: HE, Ann; BATRA, Sajeev  
; APPLICANT: LO, Terence P.; NGUYEN, Dannel B.  
; APPLICANT: BURRILL, John D.; MARCUS, Gregory A.  
; APPLICANT: ZINGLER, Kurt A.; GANDHI, Ameena R.  
; APPLICANT: LAL, Preeti G.; KEARNEY, Liam  
; APPLICANT: BURFORD, Neil; YAO, Monique G.  
; APPLICANT: CHAWLA, Narinder K.; ELLIOT, Vicki S.  
; APPLICANT: ARVIZU, Chandra S.; KHAN, Farrah A.  
; APPLICANT: BAUGHN, Mariah R.; HAFALIA, April, J.A.  
; APPLICANT: POITICKY, Jennifer L.; AU-YOUNG, Janice K.  
; APPLICANT: LU, Yan; BOROWSKY, Mark L.



APPLICANT: LU, Dyung Aina M.; RAMKUMAR, Jayalaxmi  
APPLICANT: YANG, Junming; GURURAJAN, Rajagopal  
APPLICANT: WARREN, Bridget A.; GIETZEN, Kimberly J.  
APPLICANT: XU, Yuming; KALLICK, Deborah A.  
APPLICANT: LEE, Ernestine A.; THANGAVELU, Kavitha  
APPLICANT: DELEGANE, Angelo M.; LEE, Sally  
TITLE OF INVENTION: EXTRACELLULAR MATRIX AND CELL ADHESION MOLECULES  
FILE REFERENCE: PF-0794 USN  
CURRENT APPLICATION NUMBER: US/10/312,352  
CURRENT FILING DATE: 2002-12-18  
PRIORITY APPLICATION NUMBER: PCT/US01/21067  
PRIORITY FILING DATE: 2001-06-29  
PRIORITY APPLICATION NUMBER: US 60/215,454  
PRIORITY FILING DATE: 2000-06-30  
PRIORITY APPLICATION NUMBER: US 60/219,462  
PRIORITY FILING DATE: 2000-07-18  
PRIORITY APPLICATION NUMBER: US 60/240,111  
PRIORITY FILING DATE: 2000-10-12  
PRIORITY APPLICATION NUMBER: US 60/240,106  
PRIORITY FILING DATE: 2000-10-12  
PRIORITY APPLICATION NUMBER: US 60/244,021  
PRIORITY FILING DATE: 2000-10-27  
PRIORITY APPLICATION NUMBER: US 60/248,887  
PRIORITY FILING DATE: 2000-11-14  
PRIORITY APPLICATION NUMBER: US 60/249,570  
PRIORITY FILING DATE: 2000-11-16  
NUMBER OF SEQ ID NOS: 72  
SOFTWARE: PERL Program  
SEQ ID NO 33  
LENGTH: 736  
TYPE: PRT  
ORGANISM: Homo sapiens  
FEATURE:  
NAME/KEY: misc\_feature  
OTHER INFORMATION: Incyte ID No. US20040053824A1 6780147CD1  
US-10-312-352-33

Query Match 99.4%; Score 3846; DB 12; Length 736;  
Best Local Similarity 99.8%; Pred. No. 0;  
Matches 733; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 MAVRALKLLTLLAVAAASQAQAEVESEAGWMTDPLLFAGGTAAAYARGDWPVGLSMER 60  
DB 1 MAVRALKLLTLLAVAAASQAQAEVESEAGWMTDPLLFAGGTAAAYARGDWPVGLSMER 60

QY 61 ALRSRAALRALRLRCRTQCAADFPWELDPDWSPPAQASGAGALRDLSPFGGLLRAACL 120  
DB 61 ALRSRAALRALRLRCRTQCAADFPWELDPDWSPPAQASGAGALRDLSPFGGLLRAACL 120

QY 121 RRLCLGPPAAHSLSEMELEFRKSPYNYLOVAYFKINKLEKAVAAAHTEFFVGNPEHMEQ 180  
DB 121 RRLCLGPPAAHSLSEMELEFRKSPYNYLOVAYFKINKLEKAVAAAHTEFFVGNPEHMEQ 180

QY 181 QNLDDYQTMGSKVKEADFKDLETPHMQEFLGVRLYSESOQPAVPHLEAALQEVFVAYE 240  
DB 181 QNLDDYQTMGSKVKEADFKDLETPHMQEFLGVRLYSESOQPAVPHLEAALQEVFVAYE 240

QY 241 ECRALCEGPDYDGYNYLQVAYFKINKLEKAVAAAHTEFFVGNPEHMEQ 300  
DB 241 ECRALCEGPDYDGYNYLQVAYFKINKLEKAVAAAHTEFFVGNPEHMEQ 300

QY 301 PSHYNYLOFAYNYIGNYTOAGECAKTYLLFPNDVNMQLAYAAAMLGSEHTRSGPRE 360  
DB 301 PSHYNYLOFAYNYIGNYTOAGECAKTYLLFPNDVNMQLAYAAAMLGSEHTRSGPRE 360

QY 361 SAKYEQRSLLKELLFFAYDVDFGIPFVDPDSMTPEEVPKRLQEKQKSERETAVRISOE 420  
DB 361 SAKYEQRSLLKELLFFAYDVDFGIPFVDPDSMTPEEVPKRLQEKQKSERETAVRISOE 420

QY 421 IGLNMEIETLVEKTKESLDVSRLTREGGPLYEGISLTMSKLLNGQVVMQDVISD 480  
DB 421 IGLNMEIETLVEKTKESLDVSRLTREGGPLYEGISLTMSKLLNGQVVMQDVISD 480

QY 481 HECQELQRLTNVAATSGDGYRGQTSPTPNKFKYGVTVFKALKLQBGKVPLOSAHLYYN 540  
DB 481 HECQELQRLTNVAATSGDGYRGQTSPTPNKFKYGVTVFKALKLQBGKVPLOSAHLYYN 540

QY 541 VTEKVRRTMESYFRDLDTPLYSYSHLVCRTAEEVQAEKDDSHPVHVDNCLNAETLVC 600  
DB 541 VTEKVRRTMESYFRDLDTPLYSYSHLVCRTAEEVQAEKDDSHPVHVDNCLNAETLVC 600

QY 601 VKEPPAYTFRDYSAILYLNGDFDGNFYFTELDARTVTAEVQPCGRAVGFSSGTENPHG 660  
DB 601 VKEPPAYTFRDYSAILYLNGDFDGNFYFTELDARTVTAEVQPCGRAVGFSSGTENPHG 660

QY 661 VKAVTRGORCAIALWFLDPRHSERDRVOADDLVKMLSPPEMDLSQEOFLDAQGGPPEP 720  
DB 661 VKAVTRGORCAIALWFLDPRHSERDRVOADDLVKMLSPPEMDLSQEOFLDAQGGPPEP 720

QY 721 AQESLSGSESKEPKDEL 736  
DB 721 AQESLSGSESKEPKDEL 736

RESULT 4  
US-10-257-174-25  
Sequence 25, Application US/10257174  
Publication No. US20040034194A1  
GENERAL INFORMATION:  
APPLICANT: Agarwal, Pankaj  
APPLICANT: Murdock, Paul R.  
APPLICANT: Rizvi, Safia K.  
APPLICANT: Smith, Randall F.  
APPLICANT: Xiang, Zhaoying  
TITLE OF INVENTION: NOVEL COMPOUNDS  
FILE REFERENCE: GPS0022  
CURRENT APPLICATION NUMBER: US/10/257,174  
CURRENT FILING DATE: 2002-10-10  
PRIORITY APPLICATION NUMBER: PCT/US01/11797  
PRIORITY FILING DATE: 2001-04-11  
PRIORITY APPLICATION NUMBER: 60/196,603  
PRIORITY FILING DATE: 2000-04-13  
PRIORITY APPLICATION NUMBER: 60/199,417  
PRIORITY FILING DATE: 2000-04-24  
NUMBER OF SEQ ID NOS: 48  
SOFTWARE: FastSeq for Windows Version 3.0  
SEQ ID NO 25  
LENGTH: 708  
TYPE: PRT  
ORGANISM: Homo sapiens  
US-10-257-174-25

Query Match 92.3%; Score 3572; DB 12; Length 708;  
Best Local Similarity 93.9%; Pred. No. 0;  
Matches 691; Conservative 5; Mismatches 12; Indels 28; Gaps 3;

QY 1 MAVRALKLLTLLAVAAASQAQAEVESEAGWMTDPLLFAGGTAAAYARGDWPVGLSMER 60  
DB 1 MAVRALKLLTLLAVAAASQAQAEVESEAGWMTDPLLFAGGTAAAYARGDWPVGLSMER 60

QY 61 ALRSRAALRALRLRCRTQCAADFPWELDPDWSPPAQASGAGALRDLSPFGGLLRAACL 120  
DB 61 ALRSRAALRALRLRCRTQCAADFPWELDPDWSPPAQASGAGALRDLSPFGGLLRAACL 120

QY 121 RRLCLGPPAAHSLSEMELEFRKSPYNYLOVAYFKINKLEKAVAAAHTEFFVGNPEHMEQ 180  
DB 121 RRLCLGPPAAHSLSEMELEFRKSPYNYLOVAYFKINKLEKAVAAAHTEFFVGNPEHMEQ 180

QY 181 QNLDDYQTMGSKVKEADFKDLETPHMQEFLGVRLYSESOQPAVPHLEAALQEVFVAYE 240  
DB 181 QNLDDYQTMGSKVKEADFKDLETPHMQEFLGVRLYSESOQPAVPHLEAALQEVFVAYE 240

QY 241 ECRALCEGPDYDGYNYLQVAYFKINKLEKAVAAAHTEFFVGNPEHMEQ 300  
DB 241 ECRALCEGPDYDGYNYLQVAYFKINKLEKAVAAAHTEFFVGNPEHMEQ 300

QY 301 PSHNYLQFAYNNTGNVTOGECACKTYLLFPFNDENVNQNLAAYAAAMLGEEHTRSTIGPRE 360  
DB 301 PSHNYLQFAYNNTGNVTOGECACKTYLLFPFNDENVNQNLAAYAAAMLGEEHTRSTIGPRE 360  
QY 361 SAKVROSLKELLPAYDVFGIPFVDPDSWTPPEEVIKRLQEKOKSERETAVRISQF 420  
DB 333 SAKVROSLKELLPAYDVFGIPFVDPDSWTPPEEVIKRLQEKOKSERETAVRISQF 332  
QY 421 IGLMKKEIETLVEEKRESLDVSLRTREGGPLYEGISLTWNSKLLNGYQVVMGVID 480  
DB 393 IGLMKKEIETLVEEKRESLDVSLRTREGGPLYEGISLTWNSKLLNGYQVVMGVID 452  
QY 481 HECQELQRLTNVAATSGDYGRTSPHTPNEKFGVTVFKALKLQEGKVPLOSALHYN 540  
DB 453 HECQELQRLTNVAATSGDYGRTSPHTPNEKFGVTVFKALKLQEGKVPLOSALHYN 512  
QY 541 VTEKVRIMESYFRLDTPLYFSYSHLVCRTAIEVQAEKDDSHPVHVDNCILNAETLVC 600  
DB 513 VTEKVRIMESYFRLDTPLYFSYSHLVCRTAIEVQAEKDDSHPVHVDNCILNAETLVC 572  
QY 601 VKEPPATFRDYSAILVINGDFDGCNFFYFTLDAKTVAEVQPCGRVGFSSGTENPHG 660  
DB 573 VKEPPATFRDYSAILVINGDFDGCNFFYFTLDAKTVAEVQPCGRVGFSSGTENPHG 632  
QY 661 VKAVTRGRCALALWFTLDPHRSERDRVQADDLVKMLFSPPEMDLSQEQPLDAQQGPPEP 720  
DB 633 VKAVTRGRCALALWFTLDPHRSERDRVQADDLVKMLFSPPEMDLSQEQPLDAQQGPPEP 692  
QY 721 AQESLSGSESKPDEL 736  
DB 693 AQESLSGSESKPDEL 708  
RESULT 5  
US-10-045-815-6  
; Sequence 6, Application US/10045815  
; Publication No. US20020160498A1  
; GENERAL INFORMATION:  
; APPLICANT: Wadhwa, Renu  
; APPLICANT: Sugihara, Takashi  
; TITLE OF INVENTION: TUMOR SUPPRESSOR GENE  
; FILE REFERENCE: 06501-091001  
; CURRENT APPLICATION NUMBER: US/10/045,815  
; CURRENT FILING DATE: 2001-10-26  
; PRIOR APPLICATION NUMBER: PCT/JP00/02731  
; PRIOR FILING DATE: 2000-04-26  
; PRIOR APPLICATION NUMBER: JP 11/118806  
; PRIOR FILING DATE: 1999-04-26  
; NUMBER OF SEQ ID NOS: 15  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 6  
; LENGTH: 747  
; TYPE: PRT  
; ORGANISM: Mus musculus  
US-10-045-815-6  
Query Match 85.7%; Score 3316; DB 13; Length 747;  
Best Local Similarity 88.2%; Pred. No. 2.9e-296;  
Matches 623; Conservative 33; Mismatches 48; Indels 2; Gaps 1;

QY 25 ESEAGWMTVPDILLFAEGTAAYARGDWPVGLSMERALSRAALRLRCRTQCAADFP 84  
DB 25 ESEPGWDAAPDILLYAEGTAAYSRDWPVGLVNNERALSRAALRLRCRTQCATLP 84  
QY 85 WELDPDWSPP--AQASGAGALRDLSPFGGLLRRACLRCLGPPAAHSLSEMELEFRK 142  
DB 85 WAPDLGLGPDLSQDPSQDPAALHDLRFFGAVLRAACLRCLGPPSAHLSEELDLFENK 144  
QY 143 RSPNYLQVAYFKINKLEKAAVAHAFFVGNPEHMEMQNLNDYQTMGSKVKEADFKDLET 202  
DB 145 RSPNYLQVAYFKINKLEKAAVAHAFFVGNPEHMEMQNLNDYQTMGSKVKEADFKDLEA 204

QY 203 QPHMQEFLGVLRLYSEBQEQEAVPHLEAALQBYFVVAEBCRALCEGPDYDGYNYLEVNA 262  
DB 205 KPHMEFLGVLRLYSEBQEQEAVPHLEAALQBYFVVAEBCRALCEGPDYDGYNYLDYSA 264  
QY 263 DLFOAITDHYLOVNLCKONCVTELASHSPREKPEDFLPSHYNVLOFAYNIGNVTOAGE 322  
DB 265 DLFOAITDHYLOVNLCKONCVTELASHSPREKPEDFLPSHYNVLOFAYNIGNVTOAGE 324  
QY 323 CAKTYLLFPFNDENVNQNLAAYAAAMLGEEHTRSTIGPRESAKYRQSLLEKELLPAYDV 382  
DB 325 CAKTYLLFPFNDENVNQNLAAYAAAMLGEEHTRSTIGPRESAKYRQSLLEKELLPAYDV 384  
QY 383 FGIPEVDSDSWTPPEEVIKRLQEKOKSERETAVRISQEGNLMKEIETLVEEKRESLDV 442  
DB 385 FGIPEVDSDSWTPPEEVIKRLQEKOKSERETAVRISQEGNLMKEIETLVEEKRESLDV 444  
QY 443 SRLTREGGPLYEGISLTWNSKLLNGYQVVMGVIDSDCEQELQRLTNVAATSGDYG 502  
DB 445 SRLTREGGPLYEGISLTWNSKLLNGYQVVMGVIDSDCEQELQRLTNVAATSGDYG 504  
QY 503 QTSPTPNEKFGVTVFKALKLQEGKVPLOSALHYNVTETKVRIMESYFRLDTPLYFS 562  
DB 505 QTSPTPNEKFGVTVFKALKLQEGKVPLOSALHYNVTETKVRIMESYFRLDTPLYFS 564  
QY 563 YSHLVCRTAIEVQAEKDDSHPVHVDNCILNAETLVCVKEPPATFRDYSAILVINGDF 622  
DB 565 YSHLVCRTAIEVQAEKDDSHPVHVDNCILNAETLVCVKEPPATFRDYSAILVINGDF 624  
QY 623 DGCNFFYFTLDAKTVAEVQPCGRVGFSSGTENPHGKAVTRGRCALALWFTLDPH 682  
DB 625 DGCNFFYFTLDAKTVAEVQPCGRVGFSSGTENPHGKAVTRGRCALALWFTLDPH 684  
QY 683 SERDRVQADDLVKMLFSPPEMDLSQEQPLDAQQGPPEPQAQESLSGS 728  
DB 685 SERDRVQADDLVKMLFSPPEMDLSQEQPLDAQQGPPEPQAQESLSGS 730  
RESULT 6  
US-10-045-815-8  
; Sequence 8, Application US/10045815  
; Publication No. US20020160498A1  
; GENERAL INFORMATION:  
; APPLICANT: Wadhwa, Renu  
; APPLICANT: Sugihara, Takashi  
; APPLICANT: Ohide, Akiko  
; TITLE OF INVENTION: TUMOR SUPPRESSOR GENE  
; FILE REFERENCE: 06501-091001  
; CURRENT APPLICATION NUMBER: US/10/045,815  
; CURRENT FILING DATE: 2001-10-26  
; PRIOR APPLICATION NUMBER: PCT/JP00/02731  
; PRIOR FILING DATE: 2000-04-26  
; PRIOR APPLICATION NUMBER: JP 11/118806  
; PRIOR FILING DATE: 1999-04-26  
; NUMBER OF SEQ ID NOS: 15  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 8  
; LENGTH: 542  
; TYPE: PRT  
; ORGANISM: Mus musculus  
US-10-045-815-8  
Query Match 61.6%; Score 2384; DB 13; Length 542;  
Best Local Similarity 87.9%; Pred. No. 1.7e-210;  
Matches 451; Conservative 25; Mismatches 35; Indels 2; Gaps 1;

QY 25 ESEAGWMTVPDILLFAEGTAAYARGDWPVGLSMERALSRAALRLRCRTQCAADFP 84  
DB 25 ESEPGWDAAPDILLYAEGTAAYSRDWPVGLVNNERALSRAALRLRCRTQCATLP 84  
QY 85 WELDPDWSPP--AQASGAGALRDLSPFGGLLRRACLRCLGPPAAHSLSEMELEFRK 142  
DB 85 WAPDLGLGPDLSQDPSQDPAALHDLRFFGAVLRAACLRCLGPPSAHLSEELDLFENK 144

QY 143 RSPYNYLOVAYFKINKLEKAVAAHTFFVGNBEHMEMOONLDYYOTMSGVKEADFKDLET 202  
DB 145 RSPYNYLOVAYFKINKLEKAVAAHTFFVGNBEHMEMOONLDYYOTMSGVKEADFKDLEA 204  
QY 203 QPHMOEFLRGVRLYSEBQEQEAVPHLEAALQBYFVAYEBCRALCEGPDYDYGYNLEYNA 262  
DB 205 KPHMHEFLRGVRLYSEBQEQEAVPHLEAALQBYFVAYEBCRALCEGPDYDYGYNLDYSA 264  
QY 263 DLFOAITDHYIOVLNCKQNCVTELASHPSREKPFDFLPSHYNLYOFAYYNYGNKYTOAGE 322  
DB 265 DLFOAITDHYIOVLNCKQNCVTELASHPSREKPFDFLPSHYNLYOFAYYNYGNKYTOAIE 324  
QY 323 CAKTYLLFFPNDEVNMNOLAYVAALGEEHTRISIGPRESAKERYORSLEKELLFFAYDV 382  
DB 325 CAKTYLLFFPNDEVNMNOLAYYTMLGEEBASISPRENAEYVRPNLEKELLFFAYDI 384  
QY 383 FGIPFVDPDSWTEPEVPIKRLQEKQKSERETAVRISQETGNLMKEIETLVEKTKESLDV 442  
DB 385 FGIPFVDPDSWTEPEVPIKRLQEKQKSERETAVRISQETGNLMKEIETLVEKTKESLDV 444  
QY 443 SRLTREGGPLYEGISLTWNKLLNGYQVWMDGVISDHECOELORLTNVAATSGDGYRG 502  
DB 445 SRLTREGGPLYEGISLTWNKLVNGSQVWMDGVISDDECELOLITNVAATSGDGYRG 504  
QY 503 QTSPTPNEKFGYGVTVFKALKGQGVPLQSA 535  
DB 505 QTSPTPNEKFGYGVTVLKLKGQGVPLQSA 537

RESULT 7

US-10-045-815-2  
; Sequence 2, Application US/10045815  
; Publication No. US20020160498A1  
; GENERAL INFORMATION:  
; APPLICANT: Wachwa, Renu  
; APPLICANT: Sugihara, Takashi  
; APPLICANT: Ohide, Akiko  
; TITLE OF INVENTION: TUMOR SUPPRESSOR GENE  
; FILE REFERENCE: 06501-091001  
; CURRENT APPLICATION NUMBER: US/10/045,815  
; CURRENT FILING DATE: 2001-10-26  
; PRIOR APPLICATION NUMBER: PCT/JP00/02731  
; PRIOR FILING DATE: 2000-04-26  
; PRIOR APPLICATION NUMBER: JP 11/118806  
; PRIOR FILING DATE: 1999-04-26  
; NUMBER OF SEQ ID NOS: 15  
; SOFTWARE: Fast-Seq for Windows Version 4.0  
; SEQ ID NO 2  
; LENGTH: 363  
; TYPE: PRT  
; ORGANISM: Homo sapiens

US-10-045-815-2

Query Match 49.4%; Score 1911; DB 13; Length 363;  
Best Local Similarity 100.0%; Pred. No. 3.8e-167; Indels 0; Gaps 0;  
Matches 360; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1 MAVRALKLLTLLAVVAAASQAEVSEAGWGTPTDLLFAEGTAAVARGDWPGVLSMER 60  
DB 1 MAVRALKLLTLLAVVAAASQAEVSEAGWGTPTDLLFAEGTAAVARGDWPGVLSMER 60  
QY 61 ALRSRAALRALRLRCRTQCAADFPWELDPDWSPPAQASGAGALRDLSPFGGLLRAACL 120  
DB 61 ALRSRAALRALRLRCRTQCAADFPWELDPDWSPPAQASGAGALRDLSPFGGLLRAACL 120  
QY 121 RRCGLGPPAAHSLSEMELEFRKSPYNYLOVAYFKINKLEKAVAAHTFFVGNPEHMEMQ 180  
DB 121 RRCGLGPPAAHSLSEMELEFRKSPYNYLOVAYFKINKLEKAVAAHTFFVGNPEHMEMQ 180  
QY 181 QNLDDYOTMSGVKEADFKDLETQPHMQEFLRGVRLYSEBQEQEAVPHLEAALQYFVAYE 240  
DB 181 QNLDDYOTMSGVKEADFKDLETQPHMQEFLRGVRLYSEBQEQEAVPHLEAALQYFVAYE 240

US-10-045-815-2

Query Match 49.4%; Score 1911; DB 13; Length 363;  
Best Local Similarity 100.0%; Pred. No. 3.8e-167; Indels 0; Gaps 0;  
Matches 360; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1 MAVRALKLLTLLAVVAAASQAEVSEAGWGTPTDLLFAEGTAAVARGDWPGVLSMER 60  
DB 1 MAVRALKLLTLLAVVAAASQAEVSEAGWGTPTDLLFAEGTAAVARGDWPGVLSMER 60  
QY 61 ALRSRAALRALRLRCRTQCAADFPWELDPDWSPPAQASGAGALRDLSPFGGLLRAACL 120  
DB 61 ALRSRAALRALRLRCRTQCAADFPWELDPDWSPPAQASGAGALRDLSPFGGLLRAACL 120  
QY 121 RRCGLGPPAAHSLSEMELEFRKSPYNYLOVAYFKINKLEKAVAAHTFFVGNPEHMEMQ 180  
DB 121 RRCGLGPPAAHSLSEMELEFRKSPYNYLOVAYFKINKLEKAVAAHTFFVGNPEHMEMQ 180  
QY 181 QNLDDYOTMSGVKEADFKDLETQPHMQEFLRGVRLYSEBQEQEAVPHLEAALQYFVAYE 240  
DB 181 QNLDDYOTMSGVKEADFKDLETQPHMQEFLRGVRLYSEBQEQEAVPHLEAALQYFVAYE 240

QY 241 ECRALCEGPDYDYGYNLEYNADLFQAITDHYIOVLNCKQNCVTELASHPSREKPFDFL 300  
DB 241 ECRALCEGPDYDYGYNLEYNADLFQAITDHYIOVLNCKQNCVTELASHPSREKPFDFL 300  
QY 301 PSHYNYLOFAYYNYGNKYTOAGECAKTYLLFFPNDEVNMNOLAYVAALGEEHTRISIGPRE 360  
DB 301 PSHYNYLOFAYYNYGNKYTOAGECAKTYLLFFPNDEVNMNOLAYVAALGEEHTRISIGPRE 360

RESULT 8

US-09-833-245-1251  
; Sequence 1251, Application US/09833245  
; Publication No. US20040010134A1  
; GENERAL INFORMATION:  
; APPLICANT: Human Genome Sciences, Inc.  
; TITLE OF INVENTION: Albumin Fusion Proteins  
; FILE REFERENCE: PFS46PCT  
; CURRENT APPLICATION NUMBER: US/09/833,245  
; CURRENT FILING DATE: 2001-04-12  
; PRIOR APPLICATION NUMBER: 60/223,358  
; PRIOR FILING DATE: 2000-04-12  
; PRIOR APPLICATION NUMBER: 60/256,931  
; PRIOR FILING DATE: 2000-12-21  
; PRIOR APPLICATION NUMBER: 60/199,384  
; PRIOR FILING DATE: 2000-04-25  
; NUMBER OF SEQ ID NOS: 2267  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 1251  
; LENGTH: 359  
; TYPE: PRT  
; ORGANISM: Homo sapiens

US-09-833-245-1251

Query Match 46.5%; Score 1800; DB 11; Length 359;  
Best Local Similarity 99.4%; Pred. No. 6.4e-157; Indels 0; Gaps 0;  
Matches 340; Conservative 0; Mismatches 2; Indels 0; Gaps 0;  
QY 1 MAVRALKLLTLLAVVAAASQAEVSEAGWGTPTDLLFAEGTAAVARGDWPGVLSMER 60  
DB 1 MAVRALKLLTLLAVVAAASQAEVSEAGWGTPTDLLFAEGTAAVARGDWPGVLSMER 60  
QY 61 ALRSRAALRALRLRCRTQCAADFPWELDPDWSPPAQASGAGALRDLSPFGGLLRAACL 120  
DB 61 ALRSRAALRALRLRCRTQCAADFPWELDPDWSPPAQASGAGALRDLSPFGGLLRAACL 120  
QY 121 RRCGLGPPAAHSLSEMELEFRKSPYNYLOVAYFKINKLEKAVAAHTFFVGNPEHMEMQ 180  
DB 121 RRCGLGPPAAHSLSEMELEFRKSPYNYLOVAYFKINKLEKAVAAHTFFVGNPEHMEMQ 180  
QY 181 QNLDDYOTMSGVKEADFKDLETQPHMQEFLRGVRLYSEBQEQEAVPHLEAALQYFVAYE 240  
DB 181 QNLDDYOTMSGVKEADFKDLETQPHMQEFLRGVRLYSEBQEQEAVPHLEAALQYFVAYE 240  
QY 241 ECRALCEGPDYDYGYNLEYNADLFQAITDHYIOVLNCKQNCVTELASHPSREKPFDFL 300  
DB 241 ECRALCEGPDYDYGYNLEYNADLFQAITDHYIOVLNCKQNCVTELASHPSREKPFDFL 300  
QY 301 PSHYNYLOFAYYNYGNKYTOAGECAKTYLLFFPNDEVNMNOLA 342  
DB 301 PSHYNYLOFAYYNYGNKYTOAGECAKTYLLFFPNDEVNMNOLA 342

RESULT 9

US-10-094-749-2560  
; Sequence 2560, Application US/10094749  
; Publication No. US20030219741A1  
; GENERAL INFORMATION:  
; APPLICANT: ISOGAI, TAKAO  
; APPLICANT: SUGIYAMA, TOMOYASU  
; APPLICANT: OTSUKI, TETSUJI  
; APPLICANT: WAKAMATSU, AI  
; APPLICANT: SATO, HIROYUKI  
; APPLICANT: ISHII, SHIZUKO

US-10-094-749-2560

Query Match 49.4%; Score 1911; DB 13; Length 363;  
Best Local Similarity 100.0%; Pred. No. 3.8e-167; Indels 0; Gaps 0;  
Matches 360; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1 MAVRALKLLTLLAVVAAASQAEVSEAGWGTPTDLLFAEGTAAVARGDWPGVLSMER 60  
DB 1 MAVRALKLLTLLAVVAAASQAEVSEAGWGTPTDLLFAEGTAAVARGDWPGVLSMER 60  
QY 61 ALRSRAALRALRLRCRTQCAADFPWELDPDWSPPAQASGAGALRDLSPFGGLLRAACL 120  
DB 61 ALRSRAALRALRLRCRTQCAADFPWELDPDWSPPAQASGAGALRDLSPFGGLLRAACL 120  
QY 121 RRCGLGPPAAHSLSEMELEFRKSPYNYLOVAYFKINKLEKAVAAHTFFVGNPEHMEMQ 180  
DB 121 RRCGLGPPAAHSLSEMELEFRKSPYNYLOVAYFKINKLEKAVAAHTFFVGNPEHMEMQ 180  
QY 181 QNLDDYOTMSGVKEADFKDLETQPHMQEFLRGVRLYSEBQEQEAVPHLEAALQYFVAYE 240  
DB 181 QNLDDYOTMSGVKEADFKDLETQPHMQEFLRGVRLYSEBQEQEAVPHLEAALQYFVAYE 240  
QY 241 ECRALCEGPDYDYGYNLEYNADLFQAITDHYIOVLNCKQNCVTELASHPSREKPFDFL 300  
DB 241 ECRALCEGPDYDYGYNLEYNADLFQAITDHYIOVLNCKQNCVTELASHPSREKPFDFL 300  
QY 301 PSHYNYLOFAYYNYGNKYTOAGECAKTYLLFFPNDEVNMNOLA 342  
DB 301 PSHYNYLOFAYYNYGNKYTOAGECAKTYLLFFPNDEVNMNOLA 342

RESULT 9

US-10-094-749-2560  
; Sequence 2560, Application US/10094749  
; Publication No. US20030219741A1  
; GENERAL INFORMATION:  
; APPLICANT: ISOGAI, TAKAO  
; APPLICANT: SUGIYAMA, TOMOYASU  
; APPLICANT: OTSUKI, TETSUJI  
; APPLICANT: WAKAMATSU, AI  
; APPLICANT: SATO, HIROYUKI  
; APPLICANT: ISHII, SHIZUKO

US-10-094-749-2560

Query Match 49.4%; Score 1911; DB 13; Length 363;  
Best Local Similarity 100.0%; Pred. No. 3.8e-167; Indels 0; Gaps 0;  
Matches 360; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1 MAVRALKLLTLLAVVAAASQAEVSEAGWGTPTDLLFAEGTAAVARGDWPGVLSMER 60  
DB 1 MAVRALKLLTLLAVVAAASQAEVSEAGWGTPTDLLFAEGTAAVARGDWPGVLSMER 60  
QY 61 ALRSRAALRALRLRCRTQCAADFPWELDPDWSPPAQASGAGALRDLSPFGGLLRAACL 120  
DB 61 ALRSRAALRALRLRCRTQCAADFPWELDPDWSPPAQASGAGALRDLSPFGGLLRAACL 120  
QY 121 RRCGLGPPAAHSLSEMELEFRKSPYNYLOVAYFKINKLEKAVAAHTFFVGNPEHMEMQ 180  
DB 121 RRCGLGPPAAHSLSEMELEFRKSPYNYLOVAYFKINKLEKAVAAHTFFVGNPEHMEMQ 180  
QY 181 QNLDDYOTMSGVKEADFKDLETQPHMQEFLRGVRLYSEBQEQEAVPHLEAALQYFVAYE 240  
DB 181 QNLDDYOTMSGVKEADFKDLETQPHMQEFLRGVRLYSEBQEQEAVPHLEAALQYFVAYE 240



```
; SOFTWARE: PatentIn 2.1
; SEQ ID NO 24
; LENGTH: 736
; TYPE: PR1
; ORGANISM: Homo sapiens
US-10-094-886-24

Query Match
Best Local Similarity 38.1%; Score 1473; DB 15; Length 736;
Matches 321; Conservative 116; Mismatches 266; Indels 60; Gaps 12;

QY 6 LKLLTLLAVVAASQAQEVES-----EAGGMVMTDLPFAETGTAAYARGDWPVVLSME 59
Db 2 LRLRLPLLLLLPPPPSPGPGTQLSPGAPQAPDPLLXADGLRAYAGAWAPAVALLR 61

QY 60 RALRSRAALRALRLRCKTQCAAD-----FPWEL-----DPWSPSPAQAGALRDLSPFG 111
Db 62 EALRSQAALGRVLDGASCAADFGAALPAVLIGAPEPDSGPGTQGSW-----ERQLLR 116

QY 112 GLLRRAACLRRC---LGPAAHSLSEMELE--FRKRSYNYLVQAYKINKLEKAVAA 165
Db 117 AALRRADCLTQCAARRLPGGARLRVGSALRAFRRREPYNLYRAYQLKLDLAAAA 176

QY 166 AHTFFVGNPHMEMNQNDLYQTMGVKEADFKDLETQPHMQEFLRGVRLYSEEQPOEAV 225
Db 177 AHTFFVANPHLQWREDMAKYRMSGVRPQSPRDLTPPHWAAYDTGLELLGRQEAAGLAL 236

QY 226 PHLEALOEYFVAYEECRALCEGPDYDGYNYLEYNA----DLFOALTHYIOVLACKON 281
Db 237 PRLEALQGSIAQMESCRACEGPEEQGABEEDGASQGGYEAIAAGHWIOVLQCROR 296

QY 282 CVTELASHPSREKPFDFLPSHNYLQFAYYNIQNTYTQAGECAKTYLLFPFNDVNNQNL 341
Db 297 CVGEAATPRGSPVPDFLNLRLRHEAQAQGNLSQALNVLSVLLFPFDEAAKRAL 356

QY 342 AYTAAMLGEHTSISIGRESAKYQRSLLEKELLFPAYDVGIFPVDPDSWTPVEVIK 401
Db 357 NQYQAQGEPRP-GLGPREDIQRIFLSLGEXRQLYAMEHLGTSFKDPDPWTPAALIEP 415

QY 402 RLOEKOKSERETAVRISQEIQNLKMETETLVEEKTESLDVSLRTREGGPLYEGISLTM 461
Db 416 ALREKLEDEK-----RWDHEPVKPKPLTYWKQVLLLEGVTLIQ 456

QY 462 NSKLLNGYQYVMDGVISDHECQELQRLTNVAATSG--DQYRGTSPTHTPNEKPYGVTF 519
Db 457 DSRQLNGSERAVLDGLTTPAECGVLLQLAKDAAGAGARSYGRGRRSPHTPHERFEGTLV 516

QY 520 KALKLQEGKVPLOSAHLYNTEKVRIMESYERLDTPLYFSYSHLVCRTEAEEVQAR 579
Db 517 KAAQLARAGTVGSGAKLLLEVSRVRLTQAYFSPERPLHLSTHLVCKSAIEGEGEQR 576

QY 580 KDSHPVHVNDNCILNAETLVCKEPPAYTFRDYSAILYLNGDFDGGNFYFTELDKATVTA 639
Db 577 MDLSHPVHADNCVLDPTGSCWREPPAYTYRDSYGLLYLNDDFQGGDLPTFEPNALTVA 636

QY 640 EVQPOCGRAVFGSGTENPHGVKAVTEGQRCATALMTFLDPRHSERDRVQADDLVKMLFS 699
Db 637 RVPRCGRVAFSGGVENPHGVNAVTRGRCALMHTWAPEREQEWIEAKELLQ---E 693

QY 700 PEEMDLSEQEPDQAQGGPPPPAQ-----ESLSGSESKPKDEL 736
Db 694 SQEEEEEEMSPSKDPSPPSPRRHORVQDKTRAPRVREEL 736

RESULT 11
US-10-012-697-1493
; Sequence 1493, Application US/10012697
; Publication No. US20030215803A1
; GENERAL INFORMATION:
; APPLICANT: Escobedo, Jaime
; APPLICANT: Garcia, Fabio Dominguez
; APPLICANT: Kassam, Altaf
; APPLICANT: Lamson, George
```

```
; APPLICANT: Scott, Beth
; APPLICANT: Drmanac, Radoje
; APPLICANT: Crkvenjakov, Radomir
; APPLICANT: Dickson, Mark
; APPLICANT: Drmanac, Snezana
; APPLICANT: Labat, Ivan
; APPLICANT: Leshkowitz, Dena
; APPLICANT: Kita, David
; APPLICANT: Garcia, Veronica
; APPLICANT: Jones, Lee William
; APPLICANT: Stache-Crain, Birgit
; TITLE OF INVENTION: HUMAN GENES AND GENE EXPRESSION PRODUCTS
; TITLE OF INVENTION: ISOLATED FROM HUMAN PROSTATE
; FILE REFERENCE: 2300-16252
; CURRENT APPLICATION NUMBER: US/10/012,697
; CURRENT FILING DATE: 2003-01-21
; PRIOR APPLICATION NUMBER: 60/254,648
; PRIOR FILING DATE: 2000-12-07
; PRIOR APPLICATION NUMBER: 60/275,668
; PRIOR FILING DATE: 2001-03-13
; NUMBER OF SEQ ID NOS: 1568
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 1493
; LENGTH: 282
; TYPE: PR1
; ORGANISM: Homo sapiens
US-10-012-697-1493

Query Match
Best Local Similarity 23.4%; Score 906.5; DB 15; Length 282;
Matches 172; Conservative 48; Mismatches 62; Indels 23; Gaps 4;

QY 432 VEEKTESLDVSRVREGGPLYEGISLTMNSKLINGYQYVMDGVISDHECQELQRLTN 491
Db 1 MEKRLSPKID--RDLREGGPLYENITFVNSEQLNGTQVLLDNLVLEEQCRELHVSAS 58

QY 492 VAATSGDGYRGTSPTHTPNEKPYGVTFKALKLQEGKVPLOSAHLYNTEKVRIMES 551
Db 59 GIMLVGSDGYRGTSPTHTPNEKPEGATVTLKALSGYEGRVPLKSARLFDISEKARRIVES 118

QY 552 YERLDTPLYFSYSHLVCRTEAEEVQARERKDDSHPVHVNDNCILNAETLVCKEPPAYTFRD 611
Db 119 YFNLNSTLYFSYTHMVCRITALSGQDRNDLSHPHADNCLLDPEANECWKEPPAYTFRD 178

QY 612 YSAILYLNGDFDGGNFYFTELDKATVTAEVQPCQGRAVFGSGTENPHGVKAVTRGQRC 671
Db 179 YSAILYMNDDEGGEFIFTEMDAKTVTASIKPKGRMISFSSGGENPHGVKAVTKGKCA 238

QY 672 IALMTLDRHSERDRVQADDLVKMLFSPPEMDLSQEQPLDAQOQPPPEAESLSGSESK 731
Db 239 VALWFTLDPYRELERIQAEVIAI-----LDQE-----QQKHE-----LNIN 277

QY 732 PKDEL 736
Db 278 PKDEL 282

RESULT 12
US-09-764-868-736
; Sequence 736, Application US/09764868
; Patent No. US20020168711A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
; FILE REFERENCE: PTZ32
; CURRENT APPLICATION NUMBER: US/09/764,868
; CURRENT FILING DATE: 2001-01-17
; Prior application data removed - refer to PALM or file wrapper
; NUMBER OF SEQ ID NOS: 1510
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 736
; LENGTH: 267
; TYPE: PR1
```



Db 4 VRTQVLLDNVLSEECRELHVSAGIMLVGDGYRGKTSPTHNPKEFGATVVKALKSG 63  
QY 526 QEGKVPLQSAHLYYNVTEKVRIMESYFRLDTPLYFSYSHLVCKRTAIEEVQAEKDDSHP 585  
Db 64 YEGKVPKSAFLFYDISEKARRIVESYFMLNSTLYFSYTHMVCRTALSGQDDRRNDLSHP 123  
QY 586 VHVDCILNAETLVCKEPPAYTFRDYSAILLYLNGDFDGNFYFTELDAKTVTAEVOP 643  
Db 124 IHADNCILDPANECMKPEPPAYTFRDYSAILLYMDDFEGEFIFTEMADAKTVTASIXP 181

Search completed: July 18, 2004, 09:56:56  
Job time : 687 secs

This Page Blank (uspto)



GenCore version 5.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: July 18, 2004, 09:30:43 ; Search time 57 Seconds  
(without alignments)  
1242.052 Million cell updates/sec

Title: US-10-045-815-4  
Perfect score: 3870  
Sequence: 1 MAVRAKLKLTLLAVVAAS.....PPEPAQESLSGSEKPKDEL 736

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 283366 seqs, 96191526 residues

Total number of hits satisfying chosen parameters: 283366

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : PIR 78: \*  
1: pir1: \*  
2: pir2: \*  
3: pir3: \*  
4: pir4: \*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	500.5	12.9	431	2 A56822	synaptonemal compl
2	157.5	4.1	603	2 B86608	excinnuclease ABC,
3	157.5	4.1	603	2 F72017	excinnuclease ABC,
4	144	3.7	383	2 A96704	hypothetical prote
5	141	3.6	342	2 E70463	conserved hypotet
6	125.5	3.2	1033	2 T42701	hypothetical prote
7	124	3.2	1129	2 T43674	EGL-27 protein - C
8	123.5	3.2	640	2 F75114	hypothetical prote
9	122	3.2	1529	2 T02730	RNA-directed DNA p
10	120.5	3.1	2231	1 A46147	spectrin beta chai
11	118.5	3.1	1040	2 AH1926	hypothetical prote
12	118	3.0	1082	2 H81982	hypothetical prote
13	118	3.0	1597	2 S68420	citron - mouse
14	116.5	3.0	746	2 T19409	hypothetical prote
15	116.5	3.0	754	2 B88252	protein F44G4.1 fi
16	116	3.0	864	2 B90395	purine NTPase [imp
17	116	3.0	1166	2 A39432	ATP-dependent deox
18	115.5	3.0	1992	1 S02771	myosin heavy chain
19	115	3.0	866	2 I79267	trab protein - Esc
20	114.5	3.0	613	2 A35296	secretogranin II p
21	114	2.9	1922	2 T00637	hypothetical prote
22	113	2.9	1127	2 G71274	hypothetical prote
23	113	2.9	1805	1 A64224	hypothetical prote
24	112.5	2.9	726	2 F70355	penicillin binding
25	112.5	2.9	1287	2 T42658	hypothetical prote
26	111.5	2.9	722	2 T21854	hypothetical prote
27	111	2.9	2431	1 MNWVSF	nonstructural poly
28	110.5	2.9	1116	2 T16112	hypothetical prote
29	110	2.8	859	2 H70327	DNA mismatch repai

30 110 2.8 1788 2 T29043 hypothetrical prote  
31 109.5 2.8 543 2 A38093 transformation-sen  
32 109.5 2.8 690 2 AG2419 hypothetrical prote  
33 109 2.8 1331 2 S75000 protoporphylin IX  
34 109 2.8 2157 2 A13009 peptidase synthetase  
35 109 2.8 2566 2 E98274 hypothetrical prote  
36 108.5 2.8 887 2 S70642 ubiquitin ligase N  
37 108.5 2.8 1948 2 S00485 gene 11-1 protein  
38 108 2.8 1179 2 G95144 conserved hypotet  
39 108 2.8 2663 1 S28261 centromere protein  
40 107.5 2.8 624 2 A55220 penicillin-binding  
41 107.5 2.8 714 2 H64543 fumurate reductase  
42 107.5 2.8 729 2 T50989 hypothetrical prote  
43 107.5 2.8 1965 2 T33216 hypothetrical prote  
44 107 2.8 829 2 F83905 hypothetrical prote  
45 107 2.8 1439 2 T47766 hypothetrical prote

## ALIGNMENTS

## RESULT 1

A56822  
synaptonemal complex protein SC56 - rat  
C:Species: Rattus norvegicus (Norway rat)  
C:Date: 05-Jan-1996 #sequence\_revision 05-Jan-1996 #text\_change 04-Mar-2000  
C:Accession: A56822; S20742  
R:Chen, Q.; Pearlman, R.E.; Moens, P.B.  
Biochem. Cell Biol. 70, 1030-1038, 1992  
A:Title: Isolation and characterization of a cDNA encoding a synaptonemal complex protein  
A:Reference number: A56822; MUID:93213429; PMID:1363622  
A:Accession: A56822  
A>Status: Preliminary  
A:Molecule type: mRNA  
A:Residues: 1-431 <CHE>  
A:Cross-references: EMBL:X65454; NID:G57191; PIDN:CAA45449.1; PID:G57192  
A:Experimental source: testis  
A>Note: sequence extracted from NCBI backbone (NCBIP:129004)  
C:Superfamily: rat synaptonemal complex protein SC56

Query Match 12.9%; Score 500.5; DB 2; Length 431;

Best Local Similarity 35.9%; Pred. No. 9.3e-27;  
Matches 124; Conservative 47; Mismatches 153; Indels 21; Gaps 8;

QY 46 YARGDPGVLSMERALRSRAALRLRCRTQCAADPPWELDPWSPSP---AQASGAG 102  
Db 14 VEGESWRSEARYLEALRLRLDSEAFCHANCSG--PATSQRPAPGPGDNEGDEG 71  
QY 103 ALRDLSPFGGLLRRAACLRRCCLGPPAAHSL---SEEMELEFRKSPNYLQVAYFKINKL 159  
Db 72 WARELRLEFGHVLRAACLRRCCKRTLPAPQVPYPSRQLLRDPQNLFPYQLHYAHFKANRL 131  
QY 160 EKAVAAAHTEFVGNPHEMEMQNLDYYQTMGSKVEADFKOLETOPHMQEERLGVLYSEE 219  
Db 132 EKAVAAATFLQRPKHELTARYLNYHGMLDIGESITDLEAQFYEAFLQAVLYNSG 191  
QY 220 QFQEAHPHLEAALQRYFYVAYEBCALCEGPDYDGYNYLEYNADLFOAITDHYIQVLNCK 279  
Db 192 DFRSSTEHRALADYMTVFARCLAGCEGHEQVDFK-----DFYPAIDLPAESLQCK 245  
QY 280 QNCVTELASHSPREKPFED-FLPSHYNYLQPAYNYNIGNYTOAGECAKTYLLFFPNDEVMN 338  
Db 246 VDCEANLT--PNVGQGFVDFKVFATMYHYLQPAYKLVNDVHQAAASAASNYMLFDPKDSVMQ 303  
QY 339 QNLAY---AAMLGEEHTRSGIPRESAKEYRQSRLEKELFFAY 380  
Db 304 QNLVYVYRHRARWGLEE-EDQPREAEVLYHNQISELRELLDFTH 347

## RESULT 2

B86608  
excinnuclease ABC, subunit C [imported] - Chlamydomophila pneumoniae (strain J138)  
C:Species: Chlamydomophila pneumoniae, Chlamydia pneumoniae

C>Date: 02-Mar-2001 #sequence\_revision 02-Mar-2001 #text\_change 23-Mar-2001  
C/Accession: B86608  
R:Shirai, M.; Hirakawa, H.; Kinoto, M.; Tabuchi, M.; Kishi, F.; Ouchi, K.; Shiba, T.; Ise,  
Nucleic Acids Res. 28, 2311-2314, 2000  
A/Title: Comparison of whole genome sequences of chlamydia pneumoniae J138.  
A/Reference number: A86491; MUID:20330349; PMID:10871362  
A/Accession: B86608  
A/Status: preliminary  
A/Molecule type: DNA  
A/Residues: 1-603 <STO>  
A/Cross-references: GB:BA000008; NID:g8979314; PIDN:BA99148.1; GSPDB:GN00142  
A/Experimental source: strain J138  
C/Genetics:  
A/Gene: uvrC  
C/Superfamily: excinuclease ABC chain C

Query Match 4.1%; Score 157.5; DB 2; Length 603;  
Best Local Similarity 20.9%; Pred. No. 0.0068;  
Matches 131; Conservative 81; Mismatches 223; Indels 193; Gaps 27;

QY 51 WPGVLSMERALRGRALRALRL-----RCRTQCAADFPWELDPWSPSPAQASG 100  
DB 111 WPKV-----EAIRTKAITSSORQLIFGYPVSAEACHTLL-----EVISQWFP----- 152  
QY 101 AGALRDLSPFGGLLRRAC-----LRRCLGPPAAHLSSEMELEFRKSPNYLQVAYFKI 156  
DB 153 ---LRTCSDFRALRKPFCILYDMKRLAPCVGYCTPEEYQGTLDK-----AIIFLK 201  
QY 157 NKLEKAVAAAHFTFVGNPEHMEMQNLDYYCTMGVKEADPKLETOPHMOEF----- 209  
DB 202 GKIEEVVKLEKVIQKASDNLEFQAANYRTLSLIQKAAKQVKEFHQNDALGLYR 261  
QY 210 -----RLGVRLYS--EQQPAVPHLEALQEFVA--YBECRALCEGYP 250  
DB 262 HKQRTILTLLTVRSGKLLGARHFFENAEQDQLSSFLIQYVVSQYPIPKKILTPPL 321  
QY 251 DYDGYNLYENAD-----LFQAITDHYIOVL-----NCKQNCVTELASHPSREKPFDF-- 299  
DB 322 EPTLSYV--LNAESPRLRSPKTYGKELLDLAYRNKAYAAATL---PSSLTYPQDFQN 377  
QY 300 --LPSHYNYLOFAYNYIGNYTOAGECAKTYLLFPNDENVNQNLAAYAAMLGEBHTRSIG 357  
DB 378 ILRMSQYPIRECYDNA--HMQGAHATGVIVFENNGDFPKQ-----YRTFSI-- 423  
QY 358 PRESAGEYRQSLLEKELLFPAYDVGIPFVDPDSWTPPEEVIKRLQEKOKSERETAVRI 417  
DB 424 --DSEKQNDLALLEEVL-----RRFHSLTALPDMIV-----VDGKTHYNTKKI 469  
QY 418 SOEIGNLMKEITELVEEKTES--LDVSRRLTREGGPLLYEGISLTMSKLLNGYQVRVMD 475  
DB 470 IQTNLNLTGIVVTTAKESNHSRLNKEIKFCETFP---EGFSLPPTSNLLQFFQ----- 521  
QY 476 GVISDHECQELQRLTNVAATSGDGYRGQTSBHTNEKFGYVTVFKALKLGOEGKVPLOSA 535  
DB 522 -ILRD-----EAHRFAISKH---RKRGKALFEQEKIPGIGEV----- 555  
QY 536 HLYYNTVEKVRIMESYFRDLTPLYFSYSHLVCRTAIEEVOAERKDDSHPVHVDCILNA 595  
DB 556 -----KRLQLOK-----FKSWKQVMLSSQEELEA----- 580  
QY 596 ETLVCVKEPPAYTFRDYSAILYNGDFD 623  
DB 581 -----IPGLTKKDIAVLLARQKDFN 600

RESULT 3  
F72017  
excinuclease ABC, chain C CP0921.[imported] - Chlamydothila pneumoniae (strains CWL029 a  
C/Species: Chlamydothila pneumoniae, Chlamydia pneumoniae  
C/Date: 23-Apr-1999 #sequence\_revision 23-Apr-1999 #text\_change 11-May-2000  
C/Accession: F72017; E81522  
R:Kallman, S.; Mitchell, W.; Marathe, R.; Lammel, C.; Fan, J.; Olinger, L.; Grimwood, J.;  
Nature Genet. 21, 385-389, 1999

A/Title: Comparative genomes of Chlamydia pneumoniae and C. trachomatis.  
A/Reference number: A72000; MUID:99206606; PMID:10192388  
A/Accession: F72017  
A/Status: preliminary  
A/Molecule type: DNA  
A/Residues: 1-603 <ARN>  
A/Cross-references: GB:AE001673; GB:AE001363; NID:g4377252; PIDN:AAD19078.1; PID:g4377252;  
A/Experimental source: strain CWL029  
R:Read, T.D.; Brunham, R.C.; Shen, C.; Gill, S.R.; Heidelberg, J.F.; White, O.; Hickey, I.  
C.; Dodson, R.; Gwinn, M.; Nelson, M.; DeBoy, R.; Kolonay, J.; McClarty, G.; Salzberg,  
Nucleic Acids Res. 28, 1397-1406, 2000  
A/Title: Genome sequences of Chlamydia trachomatis MoPn and Chlamydia pneumoniae AR39.  
A/Reference number: A81500; MUID:20150255; PMID:10684935  
A/Accession: E81522  
A/Status: preliminary  
A/Molecule type: DNA  
A/Residues: 1-603 <REA>  
A/Cross-references: GB:AE002251; GB:AE002161; NID:g7189833; PIDN:AAF38706.1; PID:g7189833;  
A/Experimental source: strain AR39, HL cells  
C/Genetics:  
A/Gene: uvrC; CP0921  
C/Superfamily: excinuclease ABC chain C

Query Match 4.1%; Score 157.5; DB 2; Length 603;  
Best Local Similarity 20.9%; Pred. No. 0.0068;  
Matches 131; Conservative 81; Mismatches 223; Indels 193; Gaps 27;

QY 51 WPGVLSMERALRGRALRALRL-----RCRTQCAADFPWELDPWSPSPAQASG 100  
DB 111 WPKV-----EAIRTKAITSSORQLIFGYPVSAEACHTLL-----EVISQWFP----- 152  
QY 101 AGALRDLSPFGGLLRRAC-----LRRCLGPPAAHLSSEMELEFRKSPNYLQVAYFKI 156  
DB 153 ---LRTCSDFRALRKPFCILYDMKRLAPCVGYCTPEEYQGTLDK-----AIIFLK 201  
QY 157 NKLEKAVAAAHFTFVGNPEHMEMQNLDYYCTMGVKEADPKLETOPHMOEF----- 209  
DB 202 GKIEEVVKLEKVIQKASDNLEFQAANYRTLSLIQKAAKQVKEFHQNDALGLYR 261  
QY 210 -----RLGVRLYS--EQQPAVPHLEALQEFVA--YBECRALCEGYP 250  
DB 262 HKQRTILTLLTVRSGKLLGARHFFENAEQDQLSSFLIQYVVSQYPIPKKILTPPL 321  
QY 251 DYDGYNLYENAD-----LFQAITDHYIOVL-----NCKQNCVTELASHPSREKPFDF-- 299  
DB 322 EPTLSYV--LNAESPRLRSPKTYGKELLDLAYRNKAYAAATL---PSSLTYPQDFQN 377  
QY 300 --LPSHYNYLOFAYNYIGNYTOAGECAKTYLLFPNDENVNQNLAAYAAMLGEBHTRSIG 357  
DB 378 ILRMSQYPIRECYDNA--HMQGAHATGVIVFENNGDFPKQ-----YRTFSI-- 423  
QY 358 PRESAGEYRQSLLEKELLFPAYDVGIPFVDPDSWTPPEEVIKRLQEKOKSERETAVRI 417  
DB 424 --DSEKQNDLALLEEVL-----RRFHSLTALPDMIV-----VDGKTHYNTKKI 469  
QY 418 SOEIGNLMKEITELVEEKTES--LDVSRRLTREGGPLLYEGISLTMSKLLNGYQVRVMD 475  
DB 470 IQTNLNLTGIVVTTAKESNHSRLNKEIKFCETFP---EGFSLPPTSNLLQFFQ----- 521  
QY 476 GVISDHECQELQRLTNVAATSGDGYRGQTSBHTNEKFGYVTVFKALKLGOEGKVPLOSA 535  
DB 522 -ILRD-----EAHRFAISKH---RKRGKALFEQEKIPGIGEV----- 555  
QY 536 HLYYNTVEKVRIMESYFRDLTPLYFSYSHLVCRTAIEEVOAERKDDSHPVHVDCILNA 595  
DB 556 -----KRLQLOK-----FKSWKQVMLSSQEELEA----- 580  
QY 596 ETLVCVKEPPAYTFRDYSAILYNGDFD 623  
DB 581 -----IPGLTKKDIAVLLARQKDFN 600

Db 19 KVEQSHWKVYDLCGTAFAARNYSEAIANTHFKALRANPDEPRIMNALGLAYMEAEKVKKA 78

QY 527 EGKVPQSAHLYNVTETKVRIMESYFRLDTPLYFSYSHLVCRPAIEEVAERKDDSHPV 586

Db 359 -----EMKQLYQASKDELERQKXNYQOLEQD-----LLCQLELKLKA-----SHPI 401  
Qy 587 HVD--NCILNAETIV 599  
Db 402 PEDKGKANKCDTLL 416

RESULT 7  
T43674  
EGL-27 protein - Caenorhabditis elegans  
C:Species: Caenorhabditis elegans  
C:Date: 09-Jun-2000 #sequence\_revision 09-Jun-2000 #text\_change 21-Jul-2000  
C:Accession: T43674  
R:Herman, M.A.; Ch'ng, Q.; Hettlenbach, S.M.; Ratliff, T.M.; Kenyon, C.; Herman, R.K.  
Development 126, 1055-1064, 1999  
A:Title: EGL-27 is similar to a metastasis-associated factor and controls cell polarity  
A:Reference number: 222620; MUID:99128194; PMID:9927605  
A:Accession: T43674  
A:Status: Preliminary; translated from GB/EMBL/DBJ  
A:Molecule type: mRNA  
A:Residues: 1-1129 <HER>  
A:Cross-references: EMBL:AF096618; NID:G4689162; PIDN:AAD27790.1; PID:G4689163  
C:Genetics:  
A:Gene: egl-27  
A:Map position: 2

Query Match 3.2%; Score 124; DB 2; Length 1129;  
Best Local Similarity 19.2%; Pred. No. 3.6;  
Matches 152; Conservative 97; Mismatches 270; Indels 274; Gaps 39;

Qy 130 HSLSEMELEFRKSPNYLQVAFKINKLEKAVAAHTEFVGNPEHMEQNLDDYQTM 189  
Db 89 YRLRDSVFVEVSQNEPVIAAICGFKYTKRDVVVVKLTTRYFRAD-----DIPETS 138  
Qy 190 SGVKEADFQKLETPHMQEFLGVRLY--SEEQOEAVPHLEA--ALQGYFVAYECCALCE 247  
Db 139 LNLKQERALEINPLCPQSLNRELNSQLQTPVCSLGRKCIYEVYKDVHARTVAD 198  
Qy 248 GPYDYGNY--LEYNALFOAITDHY--IQVNCQNCQVTELASHPSREKPEFDFLPSHYN 305  
Db 199 FSLNDTFFFLCHYNQDSTKLATHYAIRVGTSTFQATLPMAECVGDSDSDRDELLYREN 258  
Qy 306 YLPAYYNGNYTQAGCAKTYLL-----FFNDEVNQNLAAYAA 346  
Db 259 SIESGEE--DYIKLARCITYTLGSHMLDSQKNARSLOVSLMLDEALIIQHRSGYKI 316  
Qy 347 --MLGEBHTRSI-----GPRESAKYRQ--RSL-----EKELLFFA 379  
Db 317 DDALSELNANDIILTDDVNMOTDQAKKFAKIKQLGKNPSRIHRELLPHSREQLSVY 376  
Qy 380 YDVFQIFVPDWSW--TFEEVIPKR-----LQEKOKSERET----- 413  
Db 377 Y-----LMKTPTEATKQAAARRVNPSTSIKRPTEKVKASRPSTEVLPDPSAS 425  
Qy 414 -----AVRISQEIGN 423  
Db 426 ESDVENNGPSGRACHCYGAESKDWHHANGLLCTDCLRLHYKKGQRLQIANRPSQVPAC 485  
Qy 424 LMKEIETLVEE-----TKTESLDVSRITREGGPLYLLEGISLTWN---SKLL 466  
Db 486 LFRKRSNDEESGVRTRAGKEORRTPPSMSSETPDORRSPSTVNSGAPNLTAETPTKKL 545  
Qy 467 NGY-----QRVMDGVISDHECQELQ-----LTNVAATSGDGYRGQTSPTHTN 510  
Db 546 NGSVKRAPKPLNGVNNVKNESSEEPASPTTPPTPLTN--CLTNGH-----PESSTEN 600  
Qy 511 EKFGVTVFALKU-----GQBGKVPLOQAHLYNNVTEKVRIMESYFLDPLPYF 561  
Db 601 ----GETISKRMKVPESYDDDDDEEGKMTIDEG-----DDMPFV 637  
Qy 562 SYHLVCRTAIEEVOAB-----RKDDSHPVHVD-----NCILNAETILVCVKEPPAYT---F 609

Db 638 -LNGFKKEESVEEIKLELNGTIKKENG--VETDPTTLTCSMEARENEVC--ETPAVVSVEI 692  
Qy 610 RDSYAILYLNGDFDGGNFYFTELDAKTVTAEVQP-----QCGRVAGFSGGTNPNGHVK 662  
Db 693 RDET-----NGE-----TNSDLKD--DENVEDPDSPEDTFLGNSVEFE--TKNAMFVR 736  
Qy 663 AVTR--GORCAI-----ALWFTLDPHSEDRVQADDLVKMLFSPPEMDLSQEQ-- 709  
Db 737 SIVRCGPRCARTDLIFKIKVGVWEKSIKEKERKVHLQN--QRIQDSKVAIQNQI 794  
Qy 710 PLDAQOGPPPAQ 722  
Db 795 KKEQOQSQTPTQ 807

RESULT 8  
F75114  
Hypothetical protein PAB0498 - Pyrococcus abyssi (strain Orsay)  
C:Species: Pyrococcus abyssi  
C:Date: 20-Aug-1999 #sequence\_revision 20-Aug-1999 #text\_change 20-Aug-1999  
C:Accession: F75114  
R:Anonymous, Genoscope  
submitted to the EMBL Data Library, July 1999  
A:Description: Pyrococcus abyssi genome sequence: Insights into archaeal chromosome struc  
A:Reference number: A75001  
A:Accession: F75114  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-640 <KAW>  
A:Cross-references: GB:AJ248285; GB:AL096836; NID:G5458067; PIDN:CAB49631.1; PID:el51552;  
A:Experimental source: strain Orsay  
C:Genetics:  
A:Gene: PAB0498

Query Match 3.2%; Score 123.5; DB 2; Length 640;  
Best Local Similarity 20.3%; Pred. No. 1.7;  
Matches 97; Conservative 68; Mismatches 140; Indels 174; Gaps 24;

Qy 80 AADPWE-----LDPDWSPPAQAGAGALRDLSPFGGLLRRAACLRCLGP--- 126  
Db 129 AAIMGEIRKDVMTGMINPDGSGPV-----GGILEKASAAHQAGAKIFL 174  
Qy 127 -PAAHSLSEMELEFRKSPNY-----YLOVAYFKI 156  
Db 175 IPEGRIQVVTETSKQIGPITQITTTQRRVDVAKYQERWGLKXVIEIRDIYDAVYFTG 234  
Qy 157 NKLE-----KAVAAHTFFVGNPEHMEQNLDDYQTMGSGVKEADFKOLETQPHMQEFLG 212  
Db 235 HKLERPKGKGVIVSTGFLKYARTDYDETNNYN-----DVKO----- 273  
Qy 213 VRLYSEEQEAVPHLEALQGYFVAYECC--ALCEGPDYDGYNLYEYNADLFOAITDH 271  
Db 274 -KLVNSDVGYSTYTLKALDDAYNVLKQARDAIDQGY--YTAMS--LDFQARIMRHVDW 330  
Qy 272 YIQV-----LNCKQNCVTEL-----ASHPSREKPEFDFLPSHY 304  
Db 331 YIDVYDGRDIEDVFFEVKNEINSVEKYSNLTIRKVTMLQAVAASETRIEAEKYLKK-- 388  
Qy 305 NYLOPAYN-----IGNYTOAGCAKTYLLFFPNDVMNQNLAYYAAMLGEHTR--SIG 357  
Db 389 --AQASYNQDYWAVGNAAYAYERAKTAEWL-----AKLGEFPAKSII 431  
Qy 358 PRESAKYRQSLLEKELL--FFAYDVFG-----IPFVDPDSWTPBEVTPKLEQKOK-- 408  
Db 432 SRDSIKKAAREQLDNAKLIVMYITSMTFGQQNLNLDMLDLKG-----ERYVEDGY 482  
Qy 409 -----SEETAVR---ISOEIG-----NLMKEIETLVEEKTESLDVSRITREGG--PLL 453  
Db 483 SAALFSAMEARIRGEVILDTIGIENSVLRDKXSMKEEAKTAIG---LAQEGTGPLL 538

RESULT 9  
T02730

RNA-directed DNA polymerase homolog T914.6 - Arabidopsis thaliana  
N;Alternate names: hypothetical protein Atg28980  
C;Species: Arabidopsis thaliana (mouse-ear cress)  
C;Date: 24-Mar-1999 #sequence\_revision 24-Mar-1999 #text\_change 16-Feb-2001  
C;Accession: T02730; C84591  
R;Rounsley, S.D.; Lin, X.; Kaul, S.; Shea, T.P.; Benito, M.I.; Town, C.D.; Fujii, C.Y.; Roun-  
sley, S.D.; Kaul, S.; Rounsley, S.D.; Shea, T.P.; Benito, M.I.; Town, C.D.; Fujii, C.Y.;  
submitted to the EMBL Data Library, August 1998  
A;Description: Arabidopsis thaliana chromosome II BAC T914 genomic sequence.  
A;Reference number: Z14710  
A;Accession: T02730  
A;Status: translated from GB/EMBL/DBJ  
A;Molecule type: DNA  
A;Residues: 1-1529 <ROU>  
A;Cross-references: EMBL:AC005315; NID:g3461834; PID:g3461840  
A;Experimental source: cultivar Columbia  
R;Lin, X.; Kaul, S.; Rounsley, S.D.; Shea, T.P.; Benito, M.I.; Town, C.D.; Fujii, C.Y.;  
M.; Koo, H.; Moffat, K.S.; Cronin, L.A.; Shen, M.; VanAken, S.E.; Umayam, L.; Tallon, L.;  
eues, D.; Nierman, W.C.; White, O.; Eisen, J.A.; Salzberg, S.L.; Fraser, C.M.; Venter, J.  
Nature 402, 761-768, 1999  
A;Title: Sequence and analysis of chromosome 2 of the plant Arabidopsis thaliana.  
A;Reference number: A84420; MUID:20083487; PMID:10617197  
A;Accession: C84591  
A;Status: preliminary  
A;Molecule type: DNA  
A;Residues: 1-1529 <STO>  
A;Cross-references: GB:AE002093; NID:g3461840; PIDN:AAC33226.1; GSPDB:GN00139  
C;Genetics:  
A;Gene: Atg28980; T914.6  
A;Map position: 2  
A;Intons: 428/1; 532/3

		Query Match	3.3%; Score 122; DB 2; Length 1529;
		Best local Similarity	20.9%; Pred. No. 7.9;
		Matches 118; Conservative	85; Mismatches 198; Indels 164; Gaps 29;
QY	284	TELASHP--SREKPFEDFLPSHYNILQFAYNIGNTQAGECAKTLLPPNPDEVNNQL	341
DB	269	TEDTKTEPLMKKEPLE-ILSKSPATLKTLDGDSHTQKVPKNPTVLQNKGEVAEEN-	326
QY	342	AYYAAMLGESHTSIGPRES---AKSYRQRSLLEKELLFFA---YDVFGIPVDPDWSWT	394
DB	327	-----ENLKDEGPWHTVSPAKVGQRQPQEVNVHASPSRFDLAVEBDTADTNG	378
QY	395	PBS---VIPKLQEKKS-----ERETAVRSOEIGNLMKETELVEEKTKESL-----	440
DB	379	HEGEIVVLSQLNDNSDVGVGOTPYPRISK---SQHKVVKSSANQTKRIWVWR	435
QY	441	DVSRLTREGGPLLYEGISLTMSKLLNGYO-----RVMDGVISDHEC	483
DB	436	DNARLTP----VFKSSOMITCSILLEGKEEFCSFIYASNFVEERRILLWDIRSHHS	490
QY	484	OELQRLT-----NVAATSGDCYGROTSPHTNEKFYGVTVFKALKLGOEKVPQLSA-	535
DB	491	PLIRKPWLICGFDFNEILEGESHNDNSPYTTP---GNRDFQEI-----GLMLLEAA	541
QY	536	-----HLYYNVTEKVR---IMESYRLDTPLYFSYSHL-----VCRTAIEBVOAE	578
DB	542	TGGRKPKFKFNVLTKLPOLPVVESHWAAPLVGSTALRYRPSKKLTKLPHLRLEGKE	601
QY	579	RKDD---SHPVHVDCILNAETL-----VCVKEPPAYTFDYSAIYLINGDF-----	622
DB	602	KLGDLPKRTREAHILICEKOATTLANPSQBETIAELEKAYT--DWTHSELSEBGFLKOKSK	659
QY	623	-----DGGNFYTEL-----DAKVTVAEQPQCRAVGFSSGTENP--	658
DB	660	LHMVNVDGNNSYFHKAQAQVKMRNSIREIRGPNATLQTSEEIK-GAERFPNEFLNRQ	718
QY	659	-----HGKVAV---TRGORCAIALWFTLDRPHERDERVQADDLVKVLSPPEMDLSQEOP	710
DB	719	SGBFHHGISVEDLNILMSYRCSVT-----DONILTRE-VTGEELQKVLFA-----MP	763
QY	711	LDAQQQPPPEPAQE-----SLSGSE	729

Db 764 NNKSPGPDGYTSEFFKATWSLTGPD 788

RESULT 10

A46147

spectrin beta chain - fruit fly (*Drosophila melanogaster*)

C:Species: *Drosophila melanogaster*

C:Date: 10-Sep-1999 #sequence\_revision 10-Sep-1999 #text\_change 21-Jun-2002

C:Accession: A46147; A33657

R:Byers, T.J.; Brandin, E.; Lue, R.A.; Winograd, E.; Branton, D.

Proc. Natl. Acad. Sci. U.S.A. 89, 6187-6191, 1992

A:Title: The complete sequence of *Drosophila* beta-spectrin reveals supra-motifs comprising

A:Reference number: A46147; MUID:9233263; PMID:1631106

A:Accession: A46147

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-2291 <BYE>

A:Cross-references: GB:M92288; NID:gl57019; PIDN:AAA28399.1; PID:gl57020

A:Note: sequence extracted from NCB1 backbone (NCBIP:108607)

R:Byers, T.J.; Husain-Chishui, A.; Dubreuil, R.R.; Branton, D.; Goldstein, L.S.B.

J. Cell Biol. 109, 1633-1641, 1989

A:Title: Sequence similarity of the amino-terminal domain of *Drosophila* beta spectrin to

A:Reference number: A33657; MUID:90009037; PMID:2677025

A:Accession: A33657

A:Status: preliminary

A:Molecule type: mRNA

A:Residues: 1-800 <BY2>

A:Cross-references: GB:M92288

C:Genetics:

A:Gene: FlyBase:beta-Spec

A:Cross-references: FlyBase:FBgn0003471

C:Superfamily: spectrin beta-G chain; alpha-actinin actin-binding domain homology; pleckstrin

C:Keywords: actin binding; cytoskeleton

F:43-267/Domain: alpha-actinin actin-binding domain homology <ACT>

F:297-408/Domain: spectrin/dystrophin repeat homology <SP1>

F:417-522/Domain: spectrin/dystrophin repeat homology <SP2>

F:1698-1804/Domain: spectrin/dystrophin repeat homology <SP3>

F:2146-2257/Domain: pleckstrin repeat homology <PK>

[illegible]

Db 1456 EPIVVKTAVERPEKIKAPL-----LEROKALEKKKEAFQFCRD---VEDEK 1500  
 Qy 598 LVCVKPPPAYTFRDYSAILYNGDFDGNFYFTTDLAKTVTAEBVQCGRAVGFSSGTEN 657  
 Db 1501 LWIDEXLPVANSPDYGNLSF-----NVHVLKXKQSLATEID-----NHE 1540  
 Qy 658 PHGVKAVTRGRCALIAWFTLDRPHSRDRVQA--DDLIV-----KMLFSPBE 701  
 Db 1541 PR-INAICNKRKLIJ-----DSGHEDAKKFEALISDLTKWQELKDAIENRRKHLESE 1593  
 Qy 702 EMDLSQOPDAOQPPPEPAQESL--SGSESKPKDEL 736  
 Db 1594 KV---QYFFDAQEBSWMSQELYVMVEDRQKDEI 1626

RESULT 11  
 AH1926  
 hypothetical protein alr0963 [imported] - Nostoc sp. (strain PCC 7120)  
 C:Species: Nostoc sp. PCC 7120  
 A:Note: Nostoc sp. strain PCC 7120 is a synonym of Anabaena sp. strain PCC 7120  
 C:Date: 14-Dec-2001 #sequence\_revision 14-Dec-2001 #text\_change 09-Dec-2002  
 C:Accession: AH1926  
 R:Kaneko, T.; Nakamura, Y.; Wolk, C.P.; Kuritz, T.; Sasamoto, S.; Watanabe, A.; Iriguchi, N.; Shimpo, S.; Sugimoto, M.; Takazawa, M.; Yamada, M.; Yasuda, M.; Tabata, S. DNA Res. 8, 205-213, 2001  
 A:Title: Complete Genomic Sequence of the Filamentous Nitrogen-fixing Cyanobacterium Anabaena PCC 7120  
 A:Reference number: AB1807; MUID:21595285; PMID:11759840  
 A:Accession: AH1926  
 A>Status: preliminary  
 A:Molecule type: DNA  
 A:Residues: 1-1040 <KUR>  
 A:Cross-references: GB:BA000019; PIDN:BA072920.1; PID:g17130309; GSPDB:GNO0179  
 A:Experimental source: strain PCC 7120  
 C:Genetics:  
 A:Gene: alr0963

Query Match 3.1%; Score 118.5; DB 2; Length 1040;  
 Best Local Similarity 21.3%; Pred. No. 7.6;  
 Matches 153; Conservative 94; Mismatches 270; Indels 201; Gaps 37;

Qy 12 LLAUVAASQA-EVESE-----AGWMTPTPLLFAEGTAAYAR-----GMPGVVLSMERA 61  
 Db 213 LEAIAAAYSAALEVTRTDFPDVWMTQNNL-----GNAHLYRIILGNKGE-----NLEAA 262  
 Qy 62 LRSRAALRALRLCRTOCAADFPEWELPDWSPSPAQAGAGALRDLSFFGGLLRRAACLR 121  
 Db 263 IAAYSA--ALEVTRT---DFF---VDWMTQNNLATAYLRIILGNRGENLENAI--- 309  
 Qy 122 RLGPPAAHLSBEM--ELEFRKR--SPNYLVQAY-----FKINKLEKAVAAAHITFFV 171  
 Db 310 -----AAFSAALEVYTRTDFPKQWAGTQNNLGEAYRNRIILGNKGENLEKAIAYSA--- 360  
 Qy 172 GNPEHMEQNNLDYQYTMMSGVKEADFKDLETQPHMQSFRGLVRLYSEEQOEQVAPHLEAA 231  
 Db 361 ---ALEVYTRTDFPEQWAGTQN-----NLG-NAYSERILGNRGENLEAA 400  
 Qy 232 LOEYFVAYEBCRALCEGPDYDGYNYLEYNADLFOAITDHYIYVLNCKQNCVTELASHPS 291  
 Db 401 IAAYSAALE-----VTRTDFPEQWAGTQNNLGTAYS----- 432  
 Qy 292 REKFEDFLPSHYNLOFAYVNYIGNYTOAGEC-AKTVLLFPNDVNMQN----- 340  
 Db 433 -----ERILGNRGENLEAA---IAAYSAALEVTRT---DFPDWMTQNNLGNAYGNRI 481  
 Qy 341 -----LAYAAMLGEETRSGPRESAKYQRSL--EKELFFAYDVF--GI 385  
 Db 482 LGNRGENLEAAIAAYSAAL--EVYTRTDFPQWAGTQNNLILNGLIYQEEKQFNLAVDYDTFAQAI 540  
 Qy 386 PFVDP-----DSWTPEEVIKLEOKOKSRETAVRISQISGNLMKEIETLVEEKT 437  
 Db 541 PTVEALRGENAGDN--IGESGKGRQAEWKNLYR--MIEVCLALGKDKTEAIEYIERSKTR 598  
 Qy 438 ESLDV-----SELTRREGGLLYEGISLTMSKLLNGYQYRVVMDGVISD-----HE- 482

Db 599 YLVLELLSKADSNLENLEPEIDSNIRFAEIKULLDDETVIIQWYIFTDQCPRAFIISKHQ 658  
 Qy 483 -----COELQRLTNVAATSGDGYRGQTSPTPTNEKPYGVTFK-ALKLQGEQKVPLOSA 535  
 Db 659 IIMQASSEDLDNLKNWT-----DNYL-----QIYGEDKQKRYQLANEQLTQTL 704  
 Qy 536 HLYNVTVKVRIVESFRLDTPLFYFSYSHLVCTAIEEVCAERKDDSHPHVDNCLINA 595  
 Db 705 HL-----NQIISLPSQYKGLIVIPHRYLHUF---PLHAFVPLANNSSQBEYLFDRPHG 756  
 Qy 596 ETLVCVKPEPPAYTFRDYSAILYLN--GDFDG-----GNFYFTTDLAKTVTAEBVQ 644  
 Db 757 VSYAPSNQLLAFTQRRVERLANLELNPFSNLFAIQTNPINDLAFTDIEVETIAADFQ 814

RESULT 12  
 H81982  
 hypothetical protein NMA0631 [imported] - Neisseria meningitidis (strain Z2491 serogroup C:Species: Neisseria meningitidis  
 C:Date: 05-May-2000 #sequence\_revision 05-May-2000 #text\_change 02-Feb-2001  
 C:Accession: H81982  
 R:Parkhill, J.; Achtman, M.; James, K.D.; Bentley, S.D.; Churcher, C.; Klee, S.R.; Morell, N.; Holtroyd, S.; Jagsels, K.; Leather, S.; Moule, S.; Mungall, K.; Quail, M.A.; Rajandream, Nature 404, 502-506, 2000  
 A:Title: Complete DNA sequence of a serogroup A strain of Neisseria meningitidis Z2491.  
 A:Reference number: AB1775; MUID:20222556; PMID:10761919  
 A:Accession: H81982  
 A>Status: preliminary  
 A:Molecule type: DNA  
 A:Residues: 1-1082 <PAR>  
 A:Cross-references: GB:AL162753; GB:AL157959; NID:G7379120; PIDN:CAB83920.1; PID:g737936;  
 A:Experimental source: serogroup A, strain Z2491  
 C:Genetics:  
 A:Gene: NMA0631  
 C:Superfamily: Neisseria meningitidis hypothetical protein NMA0631

Query Match 3.0%; Score 118; DB 2; Length 1082;  
 Best Local Similarity 18.7%; Pred. No. 8.8;  
 Matches 138; Conservative 69; Mismatches 209; Indels 322; Gaps 33;

Qy 26 SEAGGMV-----TPDLLFAEGTAAYARGDWP--GVVLSMERAL-----RGRAA 67  
 Db 20 ASVGWAMVEIDEDENPICLDLGVRFVFEAEVPTKGS LAVARLARSRRLTRRAHRL 79  
 Qy 68 LRLALRLCRTOC--AADF-----PWELDP-----DMSP----- 93  
 Db 80 LGARLLKREGVLQAADFDENGLIKSLPTWQLRAAALDRKLTPEWSAVLLHLKHGR 139  
 Qy 94 --SPAQASGAGALRDLSFFGGLLRRAACLRCLGPPAAHSLSEWELEFRKSPYNIQV 151  
 Db 140 YLSQRKNEGETADKEL---GALLGVA-----DNAHALQTG---DF--RTP----- 177  
 Qy 152 AYFKINKLEKAVAAAHPTFVGNPEHMEQNNLDYQYTMMSGVKEADFKDLETQPHVQBRL 211  
 Db 178 AELALNKEK-----ESGHIRNQRG-DYSHTFSR-----KDLQAE----- 211  
 Qy 212 GVRLYSEQPOQEAHPHLEALQE----- 234  
 Db 212 -LILIFERQKEFGPHVSGGLKEGIIETLLMTQRPALSGDAVQKMLGHCTPEPAEPAKXN 270  
 Qy 235 -----YFVAYECCALCEG-----PYD----- 251  
 Db 271 TYTAERFWLTKLNNLRILEQSGSERPLTDTERTALMDEPYRKSRLTYAQAARKLGLEDTA 330  
 Qy 252 -YDGVNLYEYNAD---LFOAITDHYI-----QVLNCKQNCVTELASHPSREK 294  
 Db 331 PFKGLRYGKDNAEASTLMEMKAYHAISRALEKEGLKDKKSPNLSPLOQBIQAFSLFK 390  
 Qy 295 PPEDF-----LPSHVNLYQFAYNI-----GNVTOAGECA-- 324  
 Db 391 TDEDITGRLKRIQIOPEILEALLKHSFDKQVQISLKALRRIVPLMEQGRKRYDEA--CAEI 448

QY 325 -----KTYLLFPNDVNM-----QNLAYAAMLGBEHTSICP-----RE 360  
Db 449 YGDHYGKNTTEKTYLFPADBIKPNVLRALSOARKVINGVVRVYSGPARIHIETARE 508  
QY 361 SAKEYRORSLEKELLFFAYDVFQIPFVDPDSTPEVTPKRLQEKQKSERETAVRISQE 420  
Db 509 VGSFKDRKEIE-----KROEENKDKREAAAKFREY 540  
QY 421 IGNLMKEIETVEEKTESLDVRLTRGPGPLLYEGISLTMSKLLNGVQVVMGVISD 480  
Db 541 FPNFVG-----EPKSKDILKRLYEQQHGKCLYSKGEINLGRNLKGYVEI-----D 587  
QY 481 HECQELQLTNVAATSGDGYRGQTSPTNPNEKFGYVTFVKALKLQEGK-VPLQSAHLYX 539  
Db 588 HALP-----FRTWDDSFNN-----KVLVLSGNQKNGQTYEYF 623  
QY 540 NVTEKVRIMESYFRIDT 557  
Db 624 NGKDNQREMGEPKARVET 641  
RESULT 13  
S68420  
citron - mouse  
C:Species: Mus musculus (house mouse)  
C>Date: 15-Feb-1997 #sequence\_revision 13-Mar-1997 #text\_change 05-Nov-1999  
C:Accession: S68420  
R:Madaule, P.; Furuyashiki, T.; Reid, T.; Ishizaki, T.; Watanabe, G.; Morii, N.; Narumiya  
FEBS Lett. 377, 243-248, 1995  
A:Title: A novel partner for the GTP-bound forms of rho and rac.  
A:Reference number: S68420; MUID:96128239; PMID:8543060  
A:Accession: S68420  
A>Status: preliminary; nucleic acid sequence not shown  
A:Molecule type: mRNA  
A:Residues: 1-1597 <MAD>  
A:Cross-references: EMBL:U93904; NID:g1079733; PID:AACS2341.1; PID:g1079734  
C:Superfamily: protein kinase C zinc-binding repeat homology; pleckstrin repeat homology  
C:Keywords: alternative splicing  
F;931-979/Domain: protein kinase C zinc-binding repeat homology <K22>  
Query Match 3.0%; Score 118; DB 2; Length 1597;  
Best Local Similarity 20.5%; Pred. No. 16;  
Matches 120; Conservative 88; Mismatches 190; Indels 186; Gaps 28;  
QY 132 LSBEEMLEFRKRPYNYLOVAYFKINKLEKAVAAHTFFVGN---PEHM--EMQQLDYY 186  
Db 343 LSEQKAMINAMDSKIRLEQRIVELSEANKLAANSLSFTORNKKAQEMISELRQOKFY 402  
QY 187 QTWGKVEADFKDLET-----OPHMQEFL-----GVPLYSEEQPQE---AVPHL 228  
Db 403 ETQAGKLEAQRKLEQLEKISHQSDSKSRLLLETRLEVSLEHEEQKLELRQUTEL 462  
QY 229 EAAEQE---YFVAYEECRALCEGPDYDGYNYLEYNADL---FQAITDHYIQLN---C 278  
Db 463 QLSQREESQLTALQARAALQSOLRQAKTELBETTAZAEETQALTARHDEIQKFDAL 522  
QY 279 KQNC--VTEL-----ASHPSRKPFDF-----LPSHNYLOFAY- 311  
Db 523 RNSCTVITDLEEQNLQNTDNEALNNQNYLSKQLDEASGANDEIVQLRSEVDHLRREIT 582  
QY 312 ---YINIGNYTOAGCACTYLLFFP-----NDEVNQNLAYA--AMLGEHT--- 353  
Db 583 EREMQLTSQKQTMALKTCTMLIEQVLDLEALNDELLEKEREQWAWRSVLGDEKQSFEC 642  
QY 354 -----RSIGPRE-----SAKEYR-----QRSLEKEILLFFAYDV 382  
Db 643 RVRELQRLMDTEKQSRARADQRTESQVVELAVKEHKABITLALQALKEQKL----- 695  
QY 383 FGIPFVDPDSTWPE-----EVIKRLQEKQKSERETAVRISQEIENLMKIEIET- 430  
Db 696 -----KAESLSKLNDELKHAVLENNARSLOQKLETELKORLLEEQAKLQQQMDLQ 749  
QY 431 -----LVEEKTESLDVSLRTR-EGGPLLYEGISLTMSKLLNGVQVVMGVISDHEQC 484

Db 750 KNIHFRLTQGLQELRADLRLKLTERSDLEYQ-----LENIQVLYSHEKVMGEGTISQ----- 801  
QY 485 ELQRLTNVAATSGDGYRGQTSPTNPNEKFGYVTFVKALKLQ-----EGKVPLOSAHLYNV 541  
Db 802 -----QTK-----LIDFLOAKMDQPAKKKKVPLQYNELKAL 833  
QY 542 TEKVRIMESYFRIDTPLYFSYSHLVCRSTAEVQAEKDKDSHP 585  
Db 834 EKEKARCAE-----LEEAL--OKTRIELRSAREEA-AHRKATDHP 870  
RESULT 14  
T19409  
hypothetical protein P44G4.1 - Caenorhabditis elegans  
C:Species: Caenorhabditis elegans  
C>Date: 15-Oct-1999 #sequence\_revision 15-Oct-1999 #text\_change 29-Oct-1999  
C:Accession: T19409; T22212  
R:Sim, M.  
submitted to the EMBL Data Library, March 1996  
A:Reference number: Z19120  
A:Accession: T19409  
A>Status: preliminary; translated from GB/EMBL/DBJ  
A:Molecule type: DNA  
A:Residues: 1-746 <WIL>  
A:Cross-references: EMBL:Z70034; PIDN:CAA93858.2; GSPDB:GN00020; CESP:F44G4.1  
A:Experimental source: clone C18E9  
R:Sim, M.  
submitted to the EMBL Data Library, June 1995  
A:Reference number: Z19530  
A:Accession: T22212  
A>Status: preliminary; translated from GB/EMBL/DBJ  
A:Molecule type: DNA  
A:Residues: 1-746 <W12>  
A:Cross-references: EMBL:Z49910; PIDN:CAA90124.2; GSPDB:GN00020; CESP:F44G4.1  
C:Genetics:  
A:Gene: CESP:F44G4.1  
A:Map position: 2  
A:Introns: 26/1; 62/1; 123/2; 326/2; 343/3; 532/3; 630/3; 654/2; 690/2; 733/3  
Query Match 3.0%; Score 116.5; DB 2; Length 746;  
Best Local Similarity 19.3%; Pred. No. 6.3; Indels 161; Gaps 32;  
Matches 121; Conservative 106; Mismatches 239;  
QY 134 EMELEFRKR--SPNYLQVAYFKINKLEKAVAAHTFFVGNPEHM--EMQQLDYYQTM 189  
Db 147 EYQVELNRILTTVQKQCSIFTFDEADKL-----PEQLLGAIKPFLDYVSTI 195  
QY 190 SGV--KEADFQDLETQPHMQEFLGVRLYSEEQPQEAAPHLEAALQEFVAYEECRALCE 247  
Db 196 SGVDFRRSIFILLSNKGGGEIARITKQYESGYPREQL-RLEAFERELM----- 243  
QY 248 GPYDYGNYLEYNADLFOAITDHYIQVLNCKQNCVTELSHPSREK-----PFEDFLP 301  
Db 244 -NFSYNEKGLQMSLISNHLIDHVPFLP-LQREHVRSCVGAYLKRGRGLDVSNVDFVE 302  
QY 302 SHYNYLQFAYNIGNYTOAGCACTYLLFFPNDVNMQNLAYAAMLGEEHTRSGPRES 361  
Db 303 RVNLNSLOYPPESSKAFSSSG-CKRV-----DAKTDLEMAKIRELLS-----SWAPKKS 349  
QY 362 AKEYRQSLLEKELFFAYDVFQIPFVDPDGN-----TPEEVIKRLQEKQKSERETAVR 416  
Db 350 KLSKEESMID-----FVEEVTG--DVDEDFEQEADMPDEVDSEDFPKKKVVKKE 403  
QY 417 ISQEIENLMKIEIETLVEE-----KTKSLDSVSLRTR-EGGPLLYEGISLTMSKLL 466  
Db 404 IKQBELTDEKLQELLEKEYEASKATATKTKD--DPKHLPKSGRQKALXR--ALRKDKEAR 459  
QY 467 NGYQVVMGVISDHECOE-----LQRLTNVAATSGDGYRGQTSPTNPNEKFGYVTFVKAL 522  
Db 460 QGERAQIRDELGESAPQKEVPKTIETSMREYDATMVNEEDDEVEHDEANDEF----- 510



QY 523 KLGQEGKVPLOSAHLYN--VTEKVRIMESYFRLDTPLY-----FSYSHLV 567  
DB 511 -----AP-----YFNRETSPKVMITMTPKAKITTFKCFELOKICPNSEIFTRKNVL 557  
QY 568 CETAIEEVOAERKDDSHPHVDNCLNAETLVCKE-PPAY-----TPRDYSAILYLNG 620  
DB 558 LKTIIEQAKEREFTDLLVVEDRKKPNGIIFCHLPEGTAIFYKINSLTFTQDLKVCY--- 614  
QY 621 DFDGNGFYFTELDAKTVTAEVQPOCGRAVG--FSSGTENPHGVKAV-----TR-GORCA- 671  
DB 615 -FD--NFFM-----YCLSKSLKLFYKFGESTSHFPEVILNNTNRLGHNIAR 657  
QY 672 -IALWFTLDPR-----HSERDRV 688  
DB 658 MLACLFPHPDKFTGRRVVTFFHNRDYI 684

RESULT 15  
B88252  
protein F44G4.1 [imported] - Caenorhabditis elegans  
C:Species: Caenorhabditis elegans  
C:Date: 10-May-2001 #sequence\_revision 10-May-2001 #text\_change 10-May-2001  
C:Accession: B88252  
R:anonymous, the C. elegans Sequencing Consortium.  
Science 282, 2012-2018, 1998  
A:Title: Genome sequence of the nematode C. elegans: a platform for investigating biolog  
A:Reference number: A75000; MUID:99069613; PMID:9851916  
A:Note: see websites genome.wustl.edu/gsc/C.elegans/ and www.sanger.ac.uk/Projects/C.ele  
A:Note: published errata appeared in Science 283, 35, 1999; Science 283, 2103, 1999, and  
A:Accession: B88252  
A>Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-754 <STO>  
A:Cross-references: GB:chr\_II; PIDN:CAA90124.1; PID:g3877079; GSPDB:GN00020; CESP:F44G4.  
A:Gene: F44G4.1  
A:Map position: 2

Query Match 3.0%; Score 116.5; DB 2; Length 754;  
Best Local Similarity 19.3%; Pred. No. 6.4;  
Matches 121; Conservative 106; Mismatches 239; Indels 161; Gaps 32;

QY 134 EEMELERKR--SPYNYLVAYKINKLEKAVAAHTFFVGNPEHM--EMQQNLDDYYQTM 189  
DB 155 EYQVELNRILITVQKQCSIFIFDEADKD-----PEQLGAIKPELDYISTY 203  
QY 190 SGV--KEADPKOLETQPHMOEFLGRVLYSEEQPEAVPHLEAALQBYFYVAYECCALCE 247  
DB 204 SGVDFRRSIFILLNKGGEIARITKEQYESGYPREQL-RLEAFERELM----- 251  
QY 248 GPYDYGNYVLEYNADLFQALITHYIQVLNCKQNCYTELASHPSREK-----PFEDFLP 301  
DB 252 -NFSYNEKGLQMSLISNHLIDHFVFLPLQREHVRSCVGYLRKRGDLVSNVDVFE 310  
QY 302 SHYNLOFAYNIGNYQAGECAKTYLLFFPNDEVNQNLAHYAAMLGEBHTRSIGIPRES 361  
DB 311 RVLSLQYFPSSKAFSSG-CKRV-----DAKTDLEMAKIRPLLS-----SMAPKKS 357  
QY 362 AKEYRQSLLEKELLFPAYDVFGIPFVDPDSW-----TPEEVI PKRLQEKQKSERETAVR 416  
DB 358 K3KKEESMID----FYEEVVTG--DVDEGDFQAEADMPDEVDSEDEPPFKKKKVKYVKE 411  
QY 417 ISQIGNLMKEIETLVE-----KTESLDVSRLTR-EGGPLLYEGISLTMSKLL 466  
DB 412 IKQELLETDEKLQELLEKYASKATATKTD--DFKHLPKSQGKALKR--ALRKDKRAR 467  
QY 467 NGQVRVNDGVIDSHECQE-----LQRLTNVAATSGDGYRGQTSPHTPNKFKYGVTVFKAL 522  
DB 468 QGERAQIRDELGSAPQKQPKIESMREYDATVWNEEDEDEVEHDEANDEF----- 518  
QY 523 KLGQEGKVPLOSAHLYN--VTEKVRIMESYFRLDTPLY-----FSYSHLV 567  
DB 519 -----AP-----YFNRETSPKVMITMTPKAKITTFKCFELOKICPNSEIFTRKNVL 565

QY 568 CETAIEEVOAERKDDSHPHVDNCLNAETLVCKE-PPAY-----TPRDYSAILYLNG 620  
DB 566 LKTIIEQAKEREFTDLLVVEDRKKPNGIIFCHLPEGTAIFYKINSLTFTQDLKVCY--- 622  
QY 621 DFDGNGFYFTELDAKTVTAEVQPOCGRAVG--FSSGTENPHGVKAV-----TR-GORCA- 671  
DB 623 -FD--NFFM-----YCLSKSLKLFYKFGESTSHFPEVILNNTNRLGHNIAR 665  
QY 672 -IALWFTLDPR-----HSERDRV 688  
DB 666 MLACLFPHPDKFTGRRVVTFFHNRDYI 692

Search completed: July 18, 2004, 09:44:06  
Job time : 61 secs



GenCore version 5.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: July 18, 2004, 09:40:18 ; Search time 58 Seconds  
(without alignments)  
655.116 Million cell updates/sec

Title: US-10-045-815-4  
Perfect score: 3870  
Sequence: 1 MAVRAKLLTLLAVVAAS.....PPEPAQESLSGSEKPKDEL 736

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 389414 seqs, 51625971 residues

Total number of hits satisfying chosen parameters: 389414

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : Issued Patents AA:\*  
1: /cgn2\_6/prodata/2/iaa/5A COMB.pep.\*  
2: /cgn2\_6/prodata/2/iaa/5B COMB.pep.\*  
3: /cgn2\_6/prodata/2/iaa/6A COMB.pep.\*  
4: /cgn2\_6/prodata/2/iaa/6B COMB.pep.\*  
5: /cgn2\_6/prodata/2/iaa/6C COMB.pep.\*  
6: /cgn2\_6/prodata/2/iaa/6D COMB.pep.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	148.5	3.8	603	4	US-09-198-452A-1011
2	118	3.0	503	4	US-09-134-000C-5682
3	116	3.0	478	4	US-09-738-946-6
4	115	3.0	553	4	US-09-134-000C-4071
5	115	3.0	325	4	US-09-972-784-4
6	114	2.9	3878	4	US-09-914-259-11
7	111	2.9	2431	1	US-07-920-281C-2
8	111	2.9	2431	3	US-08-466-277-2
9	110.5	2.9	1257	1	US-08-049-783-2
10	110.5	2.9	1257	1	US-08-158-232-6
11	110.5	2.9	1257	1	US-08-304-626-6
12	110.5	2.9	1257	1	US-08-316-301A-6
13	110.5	2.9	1257	2	US-08-611-928-6
14	110.5	2.9	1257	3	US-09-173-891-6
15	110.5	2.9	1257	3	US-09-076-137-6
16	110.5	2.9	1257	5	US-09-738-363-6
17	110.5	2.9	1257	5	PCT-US92-0362A-6
18	109	2.8	927	3	US-08-895-601-6
19	108.5	2.8	451	4	US-09-134-000C-3849
20	108	2.8	657	4	US-09-284-768A-7
21	108	2.8	1596	4	US-08-978-277A-4
22	108	2.8	2662	4	US-09-595-684B-31
23	106	2.7	608	4	US-09-284-768A-4
24	105	2.7	427	4	US-09-134-000C-4847
25	105	2.7	723	4	US-09-434-408-2
26	105	2.7	1122	4	US-09-252-991A-22843
27	105	2.7	3248	1	US-08-353-700-1

#### ALIGNMENTS

RESULT 1  
US-09-198-452A-1011

; Sequence 1011: Application US/09198452A  
; Patent No. 6555294

; GENERAL INFORMATION:

; APPLICANT: Grifffais, R.

; TITLE OF INVENTION: Chlamydia pneumoniae genomic sequence and polypeptides, fragments thereof and uses thereof, in particular for the diagnosis, prevention

; TITLE OF INVENTION: and treatment of infection

; FILE REFERENCE: 9710-003-999

; CURRENT APPLICATION NUMBER: US/09/198,452A

; CURRENT FILING DATE: 1998-11-24

; NUMBER OF SEQ ID NOS: 6849

; SEQ ID NO 1011

; LENGTH: 603

; TYPE: PRT

; ORGANISM: Chlamydia pneumoniae

; US-09-198-452A-1011

Query Match 3.8%; Score 148.5; DB 4; Length 603;

Best Local Similarity 20.8%; Pred. No. 1.3e-05;

Matches 130; Conservative 79; Mismatches 229; Indels 187; Gaps 26;

28	105	2.7	3248	5	PCT-US95-16216-1	Sequence 1, Appli
29	104.5	2.7	662	4	US-09-134-000C-5682	Sequence 5682, Ap
30	104	2.7	1071	2	US-08-975-527-1	Sequence 1, Appli
31	104	2.7	1151	4	US-09-134-001C-3242	Sequence 3242, Ap
32	103.5	2.7	525	3	US-09-113-750A-35	Sequence 35, Appli
33	103.5	2.7	911	2	US-08-484-438-10	Sequence 10, Appli
34	103	2.7	1298	4	US-09-540-236-2334	Sequence 2334, Ap
35	102.5	2.6	368	4	US-09-328-352-4661	Sequence 4661, Ap
36	102.5	2.6	513	4	US-09-134-001C-4629	Sequence 4629, Ap
37	102.5	2.6	564	4	US-09-252-991A-20698	Sequence 20698, A
38	102	2.6	906	3	US-08-630-916A-48	Sequence 48, Appli
39	102	2.6	1485	4	US-09-543-681A-5367	Sequence 5367, Ap
40	101.5	2.6	599	3	US-08-556-419-22	Sequence 22, Appli
41	101.5	2.6	610	4	US-09-800-170-22	Sequence 22, Appli
42	101.5	2.6	732	4	US-09-307-143-4	Sequence 4, Appli
43	101	2.6	2482	1	US-08-328-254-6	Sequence 6, Appli
44	100.5	2.6	316	4	US-09-241-750-2	Sequence 2, Appli
45	100.5	2.6	1151	4	US-09-328-352-4744	Sequence 4744, Ap



; TITLE OF INVENTION: ENTEROCOCCUS FAECALIS FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 032796-032
; CURRENT APPLICATION NUMBER: US/09/134,000C
; PRIOR FILING DATE: 1998-08-13
; CURRENT FILING DATE: 2001-10-04
; PRIOR FILING DATE: 1997-08-15
; NUMBER OF SEQ ID NOS: 6812
; SOFTWARE: Patentin version 3.1
; SEQ ID NO 4071
; LENGTH: 553
; TYPE: PRT
; ORGANISM: Enterococcus faecalis
US-09-134-000C-4071

Query Match
Best Local Similarity 3.0%; Score 116; DB 4; Length 553;
Best Local Similarity 18.6%; Pred. No. 0.02;
Matches 73; Conservative 67; Mismatches 119; Indels 134; Gaps 20;

QY 169 FVGNPE---HMEMQNLDDYQTM-----SGVKEADF-----KDELTQPHMOER 210
DB 208 FYTGKEDLYHELKQKATDYVIDTGVKAVASRIKEDVRFGYDLPKDAFQPTNAKI- 266

QY 211 LGVRLYSEHQPOBAPVPLEAALQEFVAYECCALCEGPDYDYGNYLNYNADLFQA--- 267
DB 267 LTVELSVKEM-----APFSWDSFALIQGETKAFEGSLL 299

QY 268 -----ITDHYIOVLNCKQNCVTELASHPSREKPFEDLP SHYNYLOPAYVNYGNYTOA 320
DB 300 AQPATNEMENEFQV-KIENNGSLTADKKTGE-----TFSKLLTFEDT 342

QY 321 GECATYLLFFPNDE--VNNQNLAYAAMLGBEHTRSIGPRSAKEYRQSRSLLEKELIFF 378
DB 343 GDIGNEYIFKPTDQGIITENV--AEITNKENS-----PVKASVQIKQTVMLP----- 390

QY 379 AYDVFGIPFVDDPSWTPEVPIKRLQEKOKSRETAVRISQEIKNLMK-EIETLVEEKTK 437
DB 391 -----VAAD-----ERLEEQKAVREFRERLAQRSTTLRPFETIMV----- 427

QY 438 ESLDVSRLTREGGLLYEGISLTMSKLLNGYQVVM-DGVISD-HECCOLQRLTNVAAT 495
DB 428 -----TMIKESNQLFFE--TTINQIKDHLRLVLPFGMTVETHEADSIEVVTREPQ 478

QY 496 SGDGYRGQTS-----HTPNEKFGYTVF 519
DB 479 VSDTWENPTNPQHQAFVNVHDQNK---GVTF 508

RESULT 5
US-09-972-784-4
; Sequence 4, Application US/09972784
; Patent No. 6566088
; GENERAL INFORMATION:
; APPLICANT: McKnight, Steven L.
; APPLICANT: Bruijk, Richard K.
; TITLE OF INVENTION: Prolyl-4-Hydroxylases
; FILE REFERENCE: UTSD0871
; CURRENT APPLICATION NUMBER: US/09/972,784
; CURRENT FILING DATE: 2001-10-04
; NUMBER OF SEQ ID NOS: 5
; SOFTWARE: Patentin Ver. 2.1
; SEQ ID NO 4
; LENGTH: 325
; TYPE: PRT
; ORGANISM: Drosophila melanogaster
US-09-972-784-4

Query Match
Best Local Similarity 3.0%; Score 115; DB 4; Length 325;
Best Local Similarity 22.3%; Pred. No. 0.01;
Matches 75; Conservative 44; Mismatches 121; Indels 96; Gaps 16;

QY 433 EETKESLD-----VSRLTREGGLLYEGISLTMSKLLNGYQVVM-DGVIS 479
DB 23 EQVFRELLDKRERYEDLCRNISDMNQYGLSVDDFLGMETGLKLTINVRSMYNAGAFQ 82

QY 480 DHECOELQRLTN-----VAATSGDGYRGQ-----TSPTTNEKF-----XGVTYFKALKL 524
DB 83 DG-----QVVTNQTPDAVAGDKIRGDKIKWVGNEPGCSNVWYLTNQIDSVVYRVNTM 137

QY 525 QGEGKVPLOSAHLYNYNTEKVRIMESYFRDLTPLYESYSLVACTALIEEVOAERKDDSH 584
DB 138 KDNGLG-----NYHIRERTRAMVACY-----PGSGTHYV----- 167

QY 585 FVHDNCILNAETLVCVKPEPAYTFRDYSAILYLNGDFD-----GNGFYFTELDAKTYTAE 640
DB 168 -MHVDN-----PQDGRVITAIYILNINWDAESGGILRIPTGTTV-AD 211

QY 641 VPOCGRAVGFSSTENPHGVKAVTRGQRCALIAFWLDPHRSERDRVQADDLVKMLFSP 700
DB 212 IEPFDRLLIFFWSIRNPHVEQPAHR-TRYAITVY--FDAKEREEALIRA-----KLENSK 265

QY 701 EEMDLSEBQPLDAQOGPPE-----PAQESLSGSESKP 732
DB 266 TNNLAAQAQ-----AQAEPDSTTTPPAAPASSASLIP 298

RESULT 6
US-09-914-259-11
; Sequence 11, Application US/09914259
; Patent No. 6495336
; GENERAL INFORMATION:
; APPLICANT: Makowski, Lee
; APPLICANT: Hyman, Paul
; APPLICANT: Williams, Mark
; TITLE OF INVENTION: STAGED ASSEMBLY OF NANOSTRUCTURES
; FILE REFERENCE: 8471-010-999
; CURRENT APPLICATION NUMBER: US/09/914,259
; CURRENT FILING DATE: 2000-11-21
; NUMBER OF SEQ ID NOS: 180
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 11
; LENGTH: 3878
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-914-259-11

Query Match
Best Local Similarity 2.9%; Score 114; DB 4; Length 3878;
Best Local Similarity 19.0%; Pred. No. 0.95;
Matches 127; Conservative 95; Mismatches 258; Indels 190; Gaps 28;

QY 179 MQNLDDYQTMGYKEADFQDLETPHMOERLGRVRLYSEBQPOEAVP-----HLEALOE 234
DB 2952 LEERKAYINTISL-----KDLITKMLQR---EAEVIDSSQSHESFSDWRGELLALQ 3003

QY 235 YFVAYECCALCEGPDYDYGNYLNYNADLFQAITDHYIOVLNCKQNCVTE-----LASH 289
DB 3004 VFL--EERSVLAA-----FRTELTAQTDAVGLNCLCEQRIQEGVEYQAAM 3050

QY 290 PSREKPFEDLP SHYNYLOPAYVNYGNYT--QAGECAKTYLLFPDENVNQNLYAYAM 347
DB 3051 ECLQADARRSLLSBIOALH-AQNNGRKITLKEQSEK-----PSQELLEYNIQKQSQ 3103

QY 348 LGEHSTRSIGPRESAKEVROSLLS-----KELLFF 378
DB 3104 MLEQVELSSMKDRATELOEQLSSEKVVVAELKSELQAOTKLETTLKAOHKHLEAF 3163

QY 379 AYDVFGIPFVDDPSWTPEVPIKRLQEKOKSRETAVRISQEIKNLMKIEITLVEEKTK 438
DB 3164 RLEV-----KDKTDEVHLINDTLASEQKKSRELOWALEKEKAKLGRS-----EERDK 3211

QY 439 SLDVSRLTREGGLP--LYEGISLTMSKLLN-----GYORVWMDGVISDHCECOLO--- 487
DB 3212 ELEDLKFSLESQKORNLOLNILLEQQOKLLNESQOKIESQRMLYDAQLESEQGRNLEQV 3271

QY 488 -----RLTNVAAT-----SGDGYRGQTSPTTNEKF----- 513
DB 3272 LLESEKVIRESMSITLDRERELHAQLQSSDG--TQSRPPLPSEDLKELQKLEKHSRI 3330

QY 514 -----XGTVTF-----KALKGOBQKVPLO-SAHLYNVNTEKVRIM 549  
DB 3331 VELLNETEKYKLSLOTRQOMEXDRQVHRKTLQTEQANTGOKKHQELQSKVEDLQRL 3390  
QY 550 E-----SYFLDTPLYFSYHLVCRALIEEVOAERKODSHPHVDNCILNAETLVCVKBP 605  
DB 3391 EERKQVYKLDLEQ-RLQIGIMEFOFKOLEERERKESRRILYON-----LNEPT 3439  
QY 606 AYTF-----BDYSAILVNGDFDGNFYFTLDAKTVAEVOQCGRVAFSSGTENPHG 660  
DB 3440 TWSLTSRTRNWLQKIEGETESN-----AKLIEMN-----GGTGCNHE 3482  
QY 661 VKAVTRQRCALALWFTLDRHSDRVRQADDLVKMLFSPPEMD--LSQEQPLDAQQGP 718  
DB 3483 LEMIRQKLOQVASKLQVLPOKASERLQFETADDEDFIWOQENIDEIILQKLTGQOG-E 3541  
QY 719 BPAQESLSGS 728  
DB 3542 EPSLVSPSTS 3551

## RESULT 7

US-07-920-281C-2  
; Sequence 2, Application US/07920281C  
; Patent No. 5739026  
; GENERAL INFORMATION:  
; APPLICANT: Garoff, Henrik  
; APPLICANT: Liljestrom, Peter  
; TITLE OF INVENTION: DNA Expression Systems Based on  
; TITLE OF INVENTION: Alphaviruses  
; NUMBER OF SEQUENCES: 27  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Birch, Stewart, Kolasch & Birch  
; STREET: P.O. Box 747  
; CITY: Falls Church  
; STATE: Virginia  
; COUNTRY: USA  
; ZIP: 22040-0747  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patent In Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/07/920,281C  
; FILING DATE: 13-AUG-1992  
; CLASSIFICATION: 435  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Murphy Jr., Gerald M.  
; REGISTRATION NUMBER: 28,977  
; REFERENCE/DOCKET NUMBER: 828-103P  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 703-241-1300  
; TELEFAX: 703-241-2848  
; TELEX: 248345  
; INFORMATION FOR SEQ ID NO: 2:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 2431 amino acids  
; TYPE: amino acid  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
; US-07-920-281C-2

Query Match 2.9%; Score 111; DB 1; Length 2431;  
Best Local Similarity 20.9%; Pred. No. 0.85;  
Matches 144; Conservative 68; Mismatches 253; Indels 224; Gaps 32;  
QY 75 CRTQCAADFFWELDPDWSPPAQASG-----AGALRDLSPFGGLLRRAACLRCLGPP 127  
DB 83 CPNRSABD-PEERL-SYAKKLAASGKVLDRITAGKITD-----LQTVWATP 127  
QY 128 AAHSLSEMELEFRKSPYNYLQVAFVKINKLEKAVAAAHFVGVNPEHMEQONLDYYQ 187

DB 128 DAEPTFLCHTDVTCRTA-----AEVAVY-----QDVAVVHA-----PTSL-----YHQ 166  
QY 188 TMSGVKEADFKDLETQPHMQE-----  
DB 187 AMKGVRTAIWYGFDTTFPFMDALAGAVTYATNWADEQVLOARNIGLCAASLTEGRGLK 226  
QY 209 -----FRLGVELYSEEQPOEAVPHLEALQOEYFVAYBECALCEGPDYD 253  
DB 227 SILRKKQLKPCDVTWFSVGSSTLYTESRKLRLSWELPSVFLHKGQSFCTC-----CDTIVSCE 284  
QY 254 GNYILE-----YNADLFOAITDHYIQVLNCKQNCVTELASHPSRKPDEFLPSH--- 303  
DB 285 GYVVKKITMCPGLYKTVGYAVTYHABGFLVCK---TTDTVKGERVSFFVCTVVPSTICD 341  
QY 304 -----YNYLQFAYVNYNIGYTAQAGECAKTYLLFFPNDVNMNOLA-Y 343  
DB 342 QWTGILATDVTPEDAQKLLVGLNQRIVN-GRTOENTNTMKVYL---PIVAVAFSKWARE 398  
QY 344 YAAMLGEEHTRSIGPRESA-----KEYRQSLLEKELLFPAYDVFGIP-----FVD 389  
DB 399 YKADLDD- -KPLGVRRERSLTCCCLWAFKTRKMTMYKKP---DTQTIIVKVPSEFNSFVI 453  
QY 390 PDSWTPEEVIPKR-----LOEKOKSRETAVRISQEIGNLMKEITLVBKTESL--- 440  
DB 454 PSLWSTGLAIPVRSRIKMLAKTKRELI PVIDASSARDAEQEKEKRELELTREALPPL 513  
QY 441 -----DVSRLTREGGPLYE-----GISLTMSKLLNGYQVRVMDGVISDH 481  
DB 514 VPIAPAEUGVVDVDEELEYHAGAGVETPRSAKLVTAQPNVDLLGNVVLSPQTVLKSS 573  
QY 482 E-----CQELQRLTNVAATSG---DGYRGQTSFPHTPNEKFYGVTVFKALKLQEGKVP 531  
DB 574 KLAPVHPLAEQVKIITHNGRAGGVQVDGVRVL- -LPCGSAIPVPEFQALS----- 623  
QY 532 LQSAHLYNVNTEKVRIMESYFRDTPLYFSYHLVCRATAEE-----VOAERKDDSHPVH 587  
DB 624 -ESATVYVNEREFVNR-----KLYHIAVHGPSLNTDEENYKVRARTDAEYVFD 672  
QY 588 VDN--CILNAET--LVCVKE---PPATTF 609  
DB 673 VDKKCCVKEEASGLVLVGLTNPFPHEF 701

## RESULT 8

US-08-466-277-2  
; Sequence 2, Application US/08466277  
; Patent No. 6190666  
; GENERAL INFORMATION:  
; APPLICANT: Garoff, Henrik  
; APPLICANT: Liljestrom, Peter  
; TITLE OF INVENTION: DNA Expression Systems Based on  
; NUMBER OF SEQUENCES: 27  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Birch, Stewart, Kolasch & Birch  
; STREET: P.O. Box 747  
; CITY: Falls Church  
; STATE: Virginia  
; COUNTRY: USA  
; ZIP: 22040-0747  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patent In Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/466,277  
; FILING DATE: 06-Jun-1995  
; CLASSIFICATION: <Unknown>  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: 07/920,281  
; FILING DATE: <Unknown>

RESULT 9  
US-08-049-783-2  
; Sequence 2, Application US/08049783

Db 365 YHVDPSYN-TQSSGNISRGSNPIPIIDNNPIIICIRNSFYKAIAGSSVLVNFKGTQG 423  
Qy 529 -----KVPLOSA-----  
Db 424 YAFQAFTGGAWDHSFIESDGAPEGHKLNIYITSPGDTLDRFINVYTLISTP-----475  
Qy 567 VCRFAIEVQAE-----KDDSHPVHVDNCILNAETLVC- 600  
Db 476 -----TINELSTEKIKGPAEKGYIKNOGIMKYKPEYINGAQPVNLE-----NQOTLIFE 527  
Qy 601 --VKEPAYTFRDYSAILYNGDFDGGNFYFTELDAKTVTAEVQPCGRVGFSSGT--E 656  
Db 528 FHASKTAQYTIR-----IRYASTQGTGKYFRLDNQELQTLNIPTSHN-----GYVTGNIGE 578  
Qy 657 N-----PHGVKAVTRGQRCALWFTLDRPHSERDRVQADDLVKMLFSPPEMDLSQEOPLD 712  
Db 579 NYDLTYTIGSYTITEGNH-----TLQIQHNDKNGMVLDD---RIEFVFK---DSLQDSP-- 624  
Qy 713 AQOGPPPEAQBSLSGSESKP 732  
Db 625 -QDSPPEVHESITIFDKSSP 643

## RESULT 10

US-08-158-232-6  
; Sequence 6, Application US/08158232  
; Patent No. 5596071  
; GENERAL INFORMATION:  
; APPLICANT: Payne, Jewel  
; APPLICANT: Kennedy, M. Keith  
; APPLICANT: Randall, John Brooks  
; APPLICANT: Meier, Henry  
; APPLICANT: Uick, Heidi Jane  
; APPLICANT: Foncetrada, Luis  
; APPLICANT: Schnepf, H. Ernest  
; APPLICANT: Schwab, George E.  
; APPLICANT: Fu, Jenny  
; TITLE OF INVENTION: No. 5596071el Bacillus thuringiensis Toxins Active  
; TITLE OF INVENTION: Against Hymenopteran Pests  
; NUMBER OF SEQUENCES: 51  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: David R. Saliwanchik  
; STREET: 2421 N.W. 41st Street, Suite A-1  
; CITY: Gainesville  
; STATE: FL  
; COUNTRY: USA  
; ZIP: 32606  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patent in Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/158,232  
; FILING DATE:  
; CLASSIFICATION: 435  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 07/887,980  
; FILING DATE: 22-MAY-1992  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 07/797,645  
; FILING DATE: 25-NOV-1991  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 07/703,977  
; FILING DATE: 22-MAY-1991  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Saliwanchik, David R.  
; REGISTRATION NUMBER: 31,794  
; REFERENCE/DOCKET NUMBER: M/SCU104.C1  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 904-375-8100  
; TELEFAX: 904-372-5800

; INFORMATION FOR SEQ ID NO: 6:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 1257 amino acids  
; TYPE: amino acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
; HYPOTHETICAL: YES  
; ANTI-SENSE: NO  
; ORIGINAL SOURCE:  
; ORGANISM: Bacillus thuringiensis  
; INDIVIDUAL ISOLATE: PS33F2  
; IMMEDIATE SOURCE:  
; CLONE: E. coli NM522 (pMYC2316) B-18785  
; FEATURE:  
; NAME/KEY: Protein  
; LOCATION: 1..1257  
; US-08-158-232-6  
  
Query Match 2.9%; Score 110.5; DB 1; Length 1257;  
Best Local Similarity 18.2%; Pred. NO. 0.3;  
Matches 113; Conservative 90; Mismatches 218; Indels 199; Gaps 31;  
  
Qy 222 QEAVPHLEAALQEFYVAYEECRALCEGPDYDGYNLEYNADLFLQAITDH--YIQVLNCK 279  
Db 114 BELKQIEALIQDITNYQD--AINQKKFD-----SLQKTINLYTVAIDNNDYIAKTQL 166  
Qy 280 QNCVTELASHPSREKPFEDFLFESHYNLYQFAYINIGNYTOAGECAKTYLLFFPNDVMNQ 339  
Db 167 ENLSILTSDIS-----IFIPEGYETGGLPYAM-----VANAHILLRLDAIYNAE 212  
Qy 340 NIAY-----YAMLGEEHTRSI-----GPRESAKEYQRSLEKELFFA 379  
Db 213 KLGFSDKEVDTHKKYIKMTIHNTHEAVIKAFNLGLDKFKSLDVNSYKKNYIKGMTVM 272  
Qy 380 YDVGFI-PFVDPDSDWTPB-----EVIKRLQEKQKSERETAARIQSEIGNL 424  
Db 273 LDVALWTFDPDHYQKEVEIEFTRTISSPIYQVPKKNQNTSSIVPSDLFHYQ--GDL 330  
Qy 425 MK-EIETLVEBKTESLDVSRLTREGGPLLYEGISLTNWSKILNGYORVMDGVISDHEC 483  
Db 331 VKLEFST-----RTDNDGLAKIFTGIRNTFYKS-----PNTHT 364  
Qy 484 QELQLRTNVAATSGDVGSGTSP-----HTP-----NEKFY-----GVTVFKALKGOEG 528  
Db 365 YHVDPSYN-TQSSGNISRGSNPIPIIDNNPIIICIRNSFYKAIAGSSVLVNFKGTQG 423  
Qy 529 -----KVPLOSA-----HLYNVVTEKVRIMESYPRLDTPLYFSYHL 566  
Db 424 YAFQAFTGGAWDHSFIESDGAPEGHKLNIYITSPGDTLDRFINVYTLISTP-----475  
Qy 567 VCRFAIEVQAE-----KDDSHPVHVDNCILNAETLVC- 600  
Db 476 -----TINELSTEKIKGPAEKGYIKNOGIMKYKPEYINGAQPVNLE-----NQOTLIFE 527  
Qy 601 --VKEPAYTFRDYSAILYNGDFDGGNFYFTELDAKTVTAEVQPCGRVGFSSGT--E 656  
Db 528 FHASKTAQYTIR-----IRYASTQGTGKYFRLDNQELQTLNIPTSHN-----GYVTGNIGE 578  
Qy 657 N-----PHGVKAVTRGQRCALWFTLDRPHSERDRVQADDLVKMLFSPPEMDLSQEOPLD 712  
Db 579 NYDLTYTIGSYTITEGNH-----TLQIQHNDKNGMVLDD---RIEFVFK---DSLQDSP-- 624  
Qy 713 AQOGPPPEAQBSLSGSESKP 732  
Db 625 -QDSPPEVHESITIFDKSSP 643

## RESULT 11

US-08-304-626-6  
; Sequence 6, Application US/08304626  
; Patent No. 5616495  
; GENERAL INFORMATION:

APPLICANT: Payne, Jewel M.  
APPLICANT: Kennedy, M. Keith  
APPLICANT: Randall, John Brooks  
APPLICANT: Meier, Henry  
APPLICANT: Wick, Heidi Jane  
APPLICANT: Fonceerrada, Luis  
APPLICANT: Schnepf, Harry E.  
APPLICANT: Schwab, George E.  
TITLE OF INVENTION: No. 5616495el Bacillus thuringiensis Isolates  
TITLE OF INVENTION: Active Against Hymenopteran Pests and Genes Encoding  
TITLE OF INVENTION: Hymenopteran-Active Toxins  
NUMBER OF SEQUENCES: 39  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: David R. Saliwanchik  
STREET: 2421 N.W. 41st Street, Suite A-1  
CITY: Gainesville  
STATE: FL  
COUNTRY: USA  
ZIP: 32606  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent in Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/304,626  
FILING DATE:  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US/07/897,980  
FILING DATE:  
ATTORNEY/AGENT INFORMATION:  
NAME: Saliwanchik, David R.  
REGISTRATION NUMBER: 31,794  
REFERENCE/DOCKET NUMBER: M/SCJ 104  
TELEPHONE: 904-375-8100  
TELEFAX: 904-372-5800  
INFORMATION FOR SEQ ID NO: 6:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 1257 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
HYPOTHETICAL: YES  
ANTI-SENSE: NO  
ORIGINAL SOURCE:  
ORGANISM: Bacillus thuringiensis  
INDIVIDUAL ISOLATE: PS3P2  
CLONE: E. coli NM522(pMYC2316) B-18785  
FEATURE:  
NAME/KEY: Protein  
LOCATION: 1..1257  
US-08-304-626-6  
Query Match 2.9%; Score 110.5; DB 1; Length 1257;  
Best Local Similarity 18.2%; Pred. No. 0.3;  
Matches 113; Conservative 90; Mismatches 218; Indels 199; Gaps 31;  
QY 222 QEAVPHLEALQRYFVAYEACRALCEGPDYDGYNLYENADLFOAITH--YIQVLNCK 279  
DB 114 ELKPOLEAIQDDITNYQD--AINOKKFD-----SLQKTINLYTVVAINDDYVKTAKTL 166  
QY 280 QNCVTLEASHPSREKPFEDPLPSHYNLYQPAYYNIGNYTOAGCAKTYLLFFPNDEVMNQ 339  
DB 167 ENLSILTSIS-----IFIPEGYETGGI.PYAM-----VANAHILLRDAIVNAE 212  
QY 340 NLAY-----YAMLGEETRIS-----GPRESAKYRQSRSLLEKELLFPA 379  
DB 213 KLGSFQKVDTHKKYIKMTIHNTHTAIVKAFNLGLDKFKSLDVSNNKXANYIKGMTMV 272

380 YDVFGL-PFVDPDSWTE-----EVIKPELOEKQKSERETAIVRISQEIGNL 424  
DB 273 LDVALMPTFDPDHYQKEVEIEFTTRITSSPIYQVPKMNQMTSSSVPSDLHFHQ--GDL 330  
QY 425 MK-BIETLVEBKTESLDVSRLTREGGPLLLEGISLTWNSKLLNGYQYVVMGVDGVIDSDEC 483  
DB 331 VKLEFST-----RTDNDGLAKI FTGIRNTFYKS-----PNTHET 364  
QY 484 QELORLTNVAATSGDVRGQTSP-----HTP-----NEKFY-----GYTVFKALXKQBG 528  
DB 365 YHVDPSYN-TOSSGNISRGSSNPIDLNPNPIISTCIRNSFYKAIGSSVLVNFADGTQG 423  
QY 529 ----KVPLQSA-----HLVYNVTEKVRIMESYFRDLTPFLYFSYSHL 566  
DB 424 YAFQAPTGGAWDHSFIESDGAPEGHKLNYYTSPGDTLRDFINVTILSTP----- 475  
QY 567 VORTAIEVQAE-----KDDSHPHVDVNCILNAETLVC- 600  
DB 476 ----TINELSTEKINGFPAEXGIKNOGIMKYKPEYINGAQPVNLE----NQOTLIFE 527  
QY 601 --VKEPPAYTRDYSAIILYLANGDFDGGNFYFTELDAKTVAEVOQCGRAVGFSSGT--E 656  
DB 528 PHASKTAQYTR-----TRYASTQGTGKGYFRLDNLOELQTLNPTSHN-----GYVTGNIGE 578  
QY 657 N-----PHGVKAVTRGQRCALIAFWTLDPHRSDRVQADDLVKMLFSPPEMDLSQEQPLD 712  
DB 579 NVDLVTIGSYTITEGNH-----TLQIHNKNGMVLN-----RIEFVFK--DSLQDSP-- 624  
QY 713 AQCGPPEPAQESLSGSESKP 732  
DB 625 -QDSPPEVHESTIIIFDKSSP 643

RESULT 12  
US-08-316-301A-6  
Sequence 6, Application US/08316301A  
Patent No. 5753492  
GENERAL INFORMATION:  
APPLICANT: Schnepf, Harry E.  
APPLICANT: Schwab, George E.  
APPLICANT: Payne, Jewel M.  
APPLICANT: Harva, Kenneth E.  
APPLICANT: Fonceerrada, Luis  
TITLE OF INVENTION: No. 5753492el Nematode-Active Toxins and Genes  
TITLE OF INVENTION: Which Code Therefor  
NUMBER OF SEQUENCES: 42  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Saliwanchik & Saliwanchik  
STREET: 2421 N.W. 41st Street, Suite A-1  
CITY: Gainesville  
STATE: FL  
COUNTRY: USA  
ZIP: 32606  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent in Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/316,301A  
FILING DATE: 30-SEP-1994  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 07/871,510  
FILING DATE: 23-APR-1992  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 07/693,018  
FILING DATE: 03-MAY-1991  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 07/565,544  
FILING DATE: 10-AUG-1990

[illegible]



ANTI-SENSE: NO  
ORIGINAL SOURCE:  
ORGANISM: Bacillus thuringiensis  
INDIVIDUAL ISOLATE: PS33F2  
IMMEDIATE SOURCE:  
CLONE: E. coli NM522(pMYC2316) B-18785  
NAME/KEY: Protein  
LOCATION: 1..1257  
US-08-611-928-6

Query Match 2.9%; Score 110.5; DB 2; Length 1257;  
Best Local Similarity 18.2%; Pred. No. 0.3; Indels 199; Gaps 31;  
Matches 113; Conservative 90; Mismatches 218; Indels 199; Gaps 31;

QY 222 QEAVPHLEALQYFYVAYEECRALCEGPDYDGYNYLEYNADLFOAITH--YIQVLNCK 279  
Db 114 BELKPOEALIQDITNYQD--AINQKFD-----SLQKTINLYTVADNNDYVTAKTQL 166

QY 280 QNCVTELASHPSREKPFEDFLPSHYNLYQFAYNIGNVTQAGECAKTYLLFFPNDEVMMQ 339  
Db 167 ENLSILTSDIS-----IFPEGETGGLPYAM-----VANAHILLRLDAIVNAE 212

QY 340 NLAY-----YAAMLGEEHTRSI-----GPRSAKEYRORSLEKELFFA 379  
Db 213 KLGFSDXKVDTHKKYIKMTIHNTBAVIKAFNLGLDKFKSLDVNSYNKANYIKGTEMV 272

QY 380 YDVFGL-PFYDPSWTPE-----EVIKRLQEKQKSPRETAVRISQSIGNL 424  
Db 273 LDVLVALMPTDPHYQKEVIEFTRTSSPIQYVPEKMNQNTSSSIVPSDLFHYQ--GDL 330

QY 425 MK-EIEFLVEEKTESLDVSLRREGGPLLYEGISLTMSKLLNGYQVVMGDIVSDHEC 483  
Db 331 VKLEFST-----RTNDGLAKIFTGIRNTFYKS-----PNTHET 364

QY 484 QELQRLTNVAATSGDYGRTGTP-----HTP-----NEKFY-----GVTVFKALKLGQEG 528  
Db 365 YHVDPSN--TQSGNISRGSNPIDLNPIISTCIRNSFYKAIAGSSVNVNFKDGTQ 423

QY 529 -----KVPLOSA-----HLYNVTKEVRIMESYRLDTPLYFSYSHL 566  
Db 424 YAFQAFTGGAWDHSFESDGAPEGHKLNITYTSPGDTLDRFINVTLT-STP----- 475

QY 567 VCRTAIEVQAE-----KDDSHPVHVDNCLNAETLVC- 600  
Db 476 -----TINELTEKIKGPAEKYIKMGIMKYKPEYINGAQPVLN-----NQQLIFE 527

QY 601 --VKEPPAYTRYSAILYNGPDGNGFYFTDELDAKTVTAEVQPCGAVGFSST--E 656  
Db 528 FHASKTAQYTIIR-----IRYASTQGTGKYFRLDNQELQTLNIPITSHN-----GYVTGNIGE 578

QY 657 N-----PHGVKAVTGQRCALMFTLDPHRSERDRVQADDLVKMLFSPSEMDLSQEQPLD 712  
Db 579 NYDLTYTGSYTITEGNH-----TLQIQRNDKNGMWLD-----RIEFVPK--DSLQDSP-- 624

QY 713 AQQGPPEPAQBSLGSSESKP 732  
Db 625 -QDSPEVHESITIFDKSSP 643

RESULT 14  
US-09-173-891-6  
Sequence 6, Application US/09173891  
Patent No. 6077937  
GENERAL INFORMATION:  
APPLICANT: Payne, Jewel  
APPLICANT: Kennedy, M. Keith  
APPLICANT: Randall, John Brooks  
APPLICANT: Meier, Henry  
APPLICANT: Uick, Heidi Jane  
APPLICANT: Foncerrada, Luis  
APPLICANT: Schnepf, H. Ernest  
APPLICANT: Schwab, George E.

APPLICANT: Fu, Jenny  
TITLE OF INVENTION: NO. 6077937el Bacillus thuringiensis Toxins Active  
TITLE OF INVENTION: Against Hymenopteran Pests  
NUMBER OF SEQUENCES: 51  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: David R. Saliwanchik  
STREET: 2421 N.W. 41st Street, Suite A-1  
CITY: Gainesville  
STATE: FL  
COUNTRY: USA  
ZIP: 32606  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/173,891  
FILING DATE:  
CLASSIFICATION:  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 08/158,232  
FILING DATE:  
APPLICATION NUMBER: US 07/887,980  
FILING DATE: 22-MAY-1992  
APPLICATION NUMBER: US 07/797,645  
FILING DATE: 25-NOV-1991  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/703,977  
FILING DATE: 22-MAY-1991  
ATTORNEY/AGENT INFORMATION:  
NAME: Saliwanchik, David R.  
REGISTRATION NUMBER: 31,794  
REFERENCE/DOCKET NUMBER: M/SCU104.C1  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 904-375-8100  
TELEFAX: 904-372-5800  
INFORMATION FOR SEQ ID NO: 6:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 1257 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
HYPOTHETICAL: YES  
ANTI-SENSE: NO  
ORIGINAL SOURCE:  
ORGANISM: Bacillus thuringiensis  
INDIVIDUAL ISOLATE: PS33F2  
IMMEDIATE SOURCE:  
CLONE: E. coli NM522(pMYC2316) B-18785  
FEATURE:  
NAME/KEY: Protein  
LOCATION: 1..1257  
US-09-173-891-6

Query Match 2.9%; Score 110.5; DB 3; Length 1257;  
Best Local Similarity 18.2%; Pred. No. 0.3;  
Matches 113; Conservative 90; Mismatches 218; Indels 199; Gaps 31;

QY 222 QEAVPHLEALQYFYVAYEECRALCEGPDYDGYNYLEYNADLFOAITH--YIQVLNCK 279  
Db 114 BELKPOEALIQDITNYQD--AINQKFD-----SLQKTINLYTVADNNDYVTAKTQL 166

QY 280 QNCVTELASHPSREKPFEDFLPSHYNLYQFAYNIGNVTQAGECAKTYLLFFPNDEVMMQ 339  
Db 167 ENLSILTSDIS-----IFPEGETGGLPYAM-----VANAHILLRLDAIVNAE 212

QY 340 NLAY-----YAAMLGEEHTRSI-----GPRSAKEYRORSLEKELFFA 379  
Db 213 KLGFSDXKVDTHKKYIKMTIHNTBAVIKAFNLGLDKFKSLDVNSYNKANYIKGTEMV 272

```

QY 380 YDVGII-PFVDPDSWTPE-----EVIKRLQKQKSERETAVRISQEIIGNL 424
Db 273 LDVALWPTFDPHYQKEVEIEFTRTISSPIYQVPVKMONTSSSIVPSDLFHQ--GDL 330
QY 425 MK-EIETLVEEKTESLDSVRLTREGGPELLYEGISLTWNSKLLNGYQVRVMDGVISDHEC 483
Db 331 VKLEFST-----RTDNDGLAKIFTGIRNTFYKS-----PNTHET 364
QY 484 QELQRLTNVAATSGDGYRGQTSP-----HTP-----NEKFY-----GVTVEKALKLQGEQ 528
Db 365 YHVDPSYN-TQSSGNISRGSSNPIDLNPIIISTCIRNSFYKAIAGSSVLVNFKDGTOG 423
QY 529 ----KVPLOSA-----HLYYVTEKVRIMESYFRLDTPLFYFSYHL 566
Db 424 YAFQAAPTGGAWDHSFIESDGAPEGHKLNIIYITSPGDTLRDFINVYTLISTP----- 475
QY 567 VCRTAIEEVQAEER-----KDDSHPVHVDNCILNAETLVC- 600
Db 476 ----TINELSTEKIKGPPAEKGYIKQIMKYKPEYINGAQPVNLE-----NQOTLIFE 527
QY 601 --VKEPPAYTRDYSAIYLNGDFDGNFYFTELDAKTVTAEVQPCQGRAVGFSSGT--E 656
Db 528 PHASKTAQYTIR----IRYASTQGTGKGYFRLDNQELQTLNPTSHN-----GYVTGNIGE 578
QY 657 N----PHGVKAVTRQORCAIALWFTLDPHRSERDRVQADDLVKMLFSPPEMDLSQEQPLD 712
Db 579 NYDLTYTIGSYTITEGNH-----TLQIQHNDKNGMVLN-----RIEFVPEK--DSLQDSP-- 624
QY 713 AQOGPPPEPAQSLSGSESKP 732
Db 625 -QDSPPPEVHSTIIFDKSSP 643

```

## RESULT 15

```

US-09-076-137-6
; Sequence 6, Application US/09076137B
; Patent No. 6166195
; GENERAL INFORMATION:
; APPLICANT: Schnepf, Harry E.
; APPLICANT: Schwab, George E.
; APPLICANT: Payne, Jewel M.
; APPLICANT: Narva, Kenneth E.
; APPLICANT: Focerrada, Luis
; TITLE OF INVENTION: No. 6166195el Nematode-Active Toxins and Genes Which Code
; TITLE OF INVENTION: Therefor
; FILE REFERENCE: MA-20CCCD2
; CURRENT APPLICATION NUMBER: US/09/076,137B
; CURRENT FILING DATE: 1998-05-12
; EARLIER APPLICATION NUMBER: 08/316,301
; EARLIER FILING DATE: 1994-09-30
; NUMBER OF SEQ ID NOS: 42
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 6
; LENGTH: 1257
; TYPE: PRT
; ORGANISM: Bacillus thuringiensis
US-09-076-137-6

```

```

Query Match      2.9%; Score 110.5; DB 3; Length 1257;
Best Local Similarity 18.2%; Pred. No. 0.3;
Matches 113; Conservative 90; Mismatches 218; Indels 199; Gaps 31;

QY 222 QEAVPHLEALQYFVAYEBCRALCEGFDYDGYNVLEYNADLFOAITH--YIQVLNCK 279
Db 114 EELKPOIEALIQDITNYQD--AINQKFD-----SLQKTINLYTVAIDNNDYVTAQTQL 166
QY 280 QNCVTELASHPSREKPPEDPLPSHYNYLQFAYYIGNYTOAGSCAKTYLLFFPNDVEMNQ 339
Db 167 ENLSILTSIS-----IIFEGYETGGLPYYAM-----VANAHILLERDAIVNAE 212
QY 340 NLAY-----YAAMLCEHTRSI-----GPRESAKYRQSRSLLEKELLFFA 379
Db 213 KLGFSKDEVDTHKKYIKMTTHNTEAVIKAFNLGLDKFKSLDVNSYNNKANYIKGMTENV 272

```

Search completed: July 18, 2004, 09:45:17  
Job time : 60 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - nucleic search, using frame\_plus\_p2n model

Run on: July 18, 2004, 09:58:44 ; Search time 978 Seconds  
(without alignments)  
3197.008 Million cell updates/sec

Title: US-10-045-815-4

Perfect score: 3870

Sequence: 1 MAVRALKLLTLLAVVAAS.....PPEPAQESLSGESKPKDEL 736

Scoring table:

BLOSUM62  
Xgapop 10.0 , Xgapext 0.5  
Ygapop 10.0 , Ygapext 0.5  
Fgapop 6.0 , Fgapext 7.0  
Delop 6.0 , Delext 7.0

Searched: 3373863 seqs, 2124099041 residues

Total number of hits satisfying chosen parameters: 6747726

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:

-MODE=frame+ p2n.model -DBV=xlp  
-Q/cgr2\_1/USPTO.spool\_p/US10045815/runat\_14072004\_123007\_16812/app\_query.fasta\_1.903  
-DBN\_Geneseq\_29Jan04 -QPM=fastap -SUFFIX=std.rng -MINMATCH=0.1 -LOOPECL=0  
-LOOPEXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=blosum62 -TRANS=human40.cdi  
-LIST=45 -DOCALIGN=200 -THR SCORE=pct -THR MAX=100 -THR MIN=0 -ALIGN=15  
-MODE=LOCAL -OUTFMT=ptc -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000  
-USR=US10045815 @CGN 1 1 885 @runat\_14072004\_123007\_16812 -NCPU=6 -ICPU=3  
-NO MMAP -LARGEQUERY -NEG SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG  
-DEV TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6  
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :

- 1: Geneseq\_29Jan04.\*
- 2: Geneseq\_1990s.\*
- 3: Geneseq\_2000s.\*
- 4: Geneseq\_2001as.\*
- 5: Geneseq\_2001bs.\*
- 6: Geneseq\_2002s.\*
- 7: Geneseq\_2003as.\*
- 8: Geneseq\_2003bs.\*
- 9: Geneseq\_2003cs.\*
- 10: Geneseq\_2004s.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	3870	100.0	2600	3 AAC64725	Aac64725 Human tum
2	3861	99.8	2242	6 ABZ11346	Abz11346 Human pol
3	3846	99.4	2211	6 AAS17573	Aas17573 DNA encod
4	3846	99.4	2563	5 AAF93800	Aaf93800 Human cDN
5	3846	99.4	2585	4 AAH14589	Aah14589 Human cDN
6	3844	99.3	2524	4 AAH14513	Aah14513 Human cDN
7	3820	98.7	2577	4 AAD08489	Aad08489 Human sec
8	3801	98.2	2753	6 ABK94910	Abk94910 Human nov

9	3801	98.2	2753	6 ABK94933	Abk94933 Human nov
10	3792	98.0	2615	4 AAD08507	Aad08507 Human sec
11	3780	97.7	2829	3 AAC64724	Aac64724 Human tum
12	3726.5	96.3	2993	4 AAH16288	Aah16288 Human cDN
13	3645	94.2	2152	6 ABZ11345	Abz11345 Human pol
14	3572	92.3	2127	6 AAS17572	Aas17572 DNA encod
15	3323	85.9	2416	3 AAC64726	Aac64726 Mouse tum
16	3120	80.6	2322	3 AAC64727	Aac64727 Mouse tum
17	2148	55.5	1607	6 ABQ76716	Abq76716 Human leu
18	2147	55.5	1309	6 ABZ11344	Abz11344 Human pol
19	1751	45.2	3396	7 ADA53353	Ada53353 Human cod
20	1747	45.1	2127	9 ACF79584	Acf79584 Human pro
21	1747	45.1	2261	9 ACF79585	Acf79585 Human pro
22	1735	44.8	1595	3 AAC76943	Aac76943 Human ORF
23	1715.5	44.3	2665	7 ACA03965	Aca03965 cDNA down
24	1575.5	40.7	2789	4 AAH14459	Aah14459 Human cDN
25	1473	38.1	2305	7 ABX70413	Abx70413 DNA encod
26	1247.5	32.2	2129	8 AAL61038	Aal61038 Human pro
27	1186	30.6	2248	7 ABX34670	Abx34670 Human mdd
28	1108	28.6	703	4 AAH06792	Aah06792 Human cDN
29	1089.5	28.2	774	4 AAH06885	Aah06885 Human cDN
30	1072	27.7	682	4 AAH196734	Aah196734 Human neu
31	1040	26.9	652	5 AAF93998	Aaf93998 Primer sp
32	1029.5	26.6	738	4 AAH06612	Aah06612 Human cDN
33	977	25.2	550	7 ACA57060	Aca57060 Human adi
34	906.5	23.4	849	8 ADB83188	Adb83188 Human cDN
35	836.5	21.6	1023	4 AAS27088	Aas27088 cDNA enco
36	836.5	21.6	1023	4 ABK43318	Abk43318 DNA encod
37	836.5	21.6	1023	9 ADB93266	Adb93266 Human cDN
38	807.5	20.9	793	9 ACF79586	Acf79586 Human pro
39	804	20.8	1849	4 ABK43627	Abk43627 DNA encod
40	790	20.4	681	5 AAS80583	Aas80583 DNA encod
41	710.5	18.4	685	9 ACF79590	Acf79590 Human pro
42	708.5	18.3	684	9 ACF79589	Acf79589 Human pro
43	674	17.4	800	4 AAH06448	Aah06448 Human cDN
44	674	17.4	800	9 ACF79587	Acf79587 Human pro
45	664.5	17.2	649	9 ACF79593	Acf79593 Human pro

ALIGNMENTS

RESULT 1

AAC64725

ID AAC64725 standard; cDNA; 2600 BP.

XX

AC AAC64725;

XX

DT 27-FEB-2001 (first entry)

XX

DE Human tumour suppressor Gros1-S encoding cDNA SEQ ID NO:3.

XX

KW Tumour suppressor; Gros1-L; Gros1-S; cell proliferation; regulation; cancer; cytostatic; gene therapy; ss.

XX

OS Homo sapiens.

XX

PN WO200065047-A1.

XX

PD 02-NOV-2000.

XX

PF 26-APR-2000; 2000WO-JP002731.

XX

PR 26-APR-1999; 99JP-00118806.

XX

PA (CHUG-) CHUGAI RES INST MOLECULAR MEDICINE INC.

XX

PI Wadhwa R, Sugihara T, Yoshida A;

XX

DR WPI: 2000-687340/67.

XX

DR P-PSDB; AAB36392.

XX

PT Full-length tumor suppressor genes encoding Gros1-L, Gros1-S and mouse homologs participating in regulation of cell proliferation, useful in

PT development of preventives and remedies of cancer.

XX Claim 1; Page 70-77; 114pp; Japanese.

XX The present sequence encodes the human tumour suppressor designated Gros1  
 CC -S. Gros1-L and Gros1-S have cytostatic activity and can be used in gene  
 CC therapy. Gros1-L and Gros1-S genes are useful in the development of drugs  
 CC used to treat and prevent cancer

XX SQ Sequence 2600 BP; 608 A; 757 C; 722 G; 513 T; 0 U; 0 Other;

#### Alignment Scores:

Pred. No.: 0 Length: 2600  
 Score: 3870.00 Matches: 736  
 Percent Similarity: 100.00% Conservative: 0  
 Best Local Similarity: 100.00% Mismatches: 0  
 Query Match: 100.00% Indels: 0  
 DB: 3 Gaps: 0

US-10-045-815-4 (1-736) x AAC64725 (1-2600)

QY 1 MetAlaValArgAlaLeuLysLeuThrThrLeuLeuAlaValAlaAlaSer 20  
 DB 52 ATGGCGGTACGCGGTGAAGCTGCTGACCACACTGCTGGCTGCGTGGCGCTGCTCC 111  
 QY 21 GlnAlaGluValGluSerGluAlaGlyTrpGlyMetValThrProAspLeuLeuPheAla 40  
 DB 112 CAAGCCGAGGTGCGATCCGAGGAGGATGGGCATGCTGACGCTGATCTGCTTCGCC 171  
 QY 41 GluGlyThrAlaAlaValAlaArgGlyAspTrpProGlyValValLeuSerMetGluArg 60  
 DB 172 GAGGGACCGCAGCTACGCGCGGGGAGCTGGCCGGGTGGTCTGACATGGAACCG 231  
 QY 61 AlaLeuArgSerArgAlaAlaLeuArgAlaLeuArgLeuArgCysArgThrGlnCysAla 80  
 DB 232 CGCTGGCTCCCGGGGAGCCCTCCGGCCCTTCCGCTGGCTGGCCGACCTGCTGCC 291  
 QY 81 AlaAspPheProTrpGluLeuAspProAspTrpSerProSerProAlaGlnAlaSerGly 100  
 DB 292 GCCGACTTCCGCTGGAGCTGAGCCCGAGCTGGTCCCGCCGCGGCCAGGCTCGGGC 351  
 QY 101 AlaGlyAlaLeuArgAspLeuSerPhePheGlyClyLeuLeuArgAlaAlaCysLeu 120  
 DB 352 GCCGGCCCTTGGCGACCTGAGCTTCTCGGGGCTTCTGGCTGGCTGGCTGGCTGG 411  
 QY 121 ArgArgCysLeuGlyProProAlaAlaHisSerLeuSerGluGluMetGluLeuGluPhe 140  
 DB 412 CGCGCTGCTCGCGCGCGCGCGCCACTCGCTCAGCGAGAGATGGAGCTGGAGTTC 471  
 QY 141 ArgLysArgSerProTrpGlnThrMetSerGlyValAlaValAlaValAlaValAlaVal 160  
 DB 472 CGCAAGCGGAGCCCTCAACTACTCGAGGTGGCTTCAAGATCAACAAGTTGGAG 531  
 QY 161 LysAlaValAlaAlaHisThrPhePheValGlyAsnProGluHisMetGluMetGln 180  
 DB 532 AAAGCTGTGTCGAGCACACACTTCTTCTGGGCAATCTCGAGCACATGGAAATGCAG 591  
 QY 181 GlnAsnLeuAspTrpGlnThrMetSerGlyValLysGluAlaAspPheLysAspLeu 200  
 DB 592 CAGAACCTAGACTATTACCAACCATCTCGAGGTGAAGAGGCCGACTTCAAGGATCTT 651  
 QY 201 GluThrGlnProHisMetGlnGluPheArgLeuGlyValArgLeuTrpSerGluGluGln 220  
 DB 652 GAGACTCAACCCCATATGCAAGATTTTCGATGGAGTGGCTGACTCTACTCAGAGGAACAG 711  
 QY 221 ProGlnGluAlaValProHisLeuGluAlaAlaLeuGlnGluThrPheValAlaValGlu 240  
 DB 712 CCACAGAAAGGTGTGGCCCACTAGAGCGGCGCTGCAAGAAATACCTTTGGGCTATGAG 771  
 QY 241 GluCysArgAlaLeuCysGluGlyProTrpAspTrpAspGlyTrpAsnTrpLeuGluTrp 260  
 DB 772 GAGTGGCTGCTTCCGAGAGGGCCCTATGATACATGATGGCTACAACTACCTTGGATAC 831

QY 261 AsnAlaAspLeuPheGlnAlaIleThrAspHisThrIleGlnValLeuAsnCysLysGln 280  
 DB 832 AACGCTGACCTTCCAGGCCATCAGATCATTTACATCCAGGTCTCACTGTAAGCAG 891  
 QY 281 AsnCysValThrGluLeuAlaSerHisProSerArgGluLysProPheGluAspPheLeu 300  
 DB 892 AACTGTGTACGGAGCTTGTCTCCACCAAGTCGAGAGAGCCCTTTGAAGACTCTCTC 951  
 QY 301 ProSerHisTrpAsnTrpLeuGlnPheAlaTrpTrpAsnIleGlyAsnTrpThrGlnAla 320  
 DB 952 CCATCGCATTAATATCTGAGTTGCTTACTATACATTTGGGAATATTACCAAGCT 1011  
 QY 321 GlyGluCysAlaLysThrTrpLeuLeuPhePheProAsnAspGluValMetAsnGlnAsn 340  
 DB 1012 GGTGAATGTGCCAAGACCTATCTTCTTCTCCCAATGACGAGGTGATGAACCAAAAT 1071  
 QY 341 LeuAlaTrpTrpAlaAlaMetLeuGlyGluGluHisThrArgSerIleGlyProArgGlu 360  
 DB 1072 TTGGCTATTATGACGCTATCTTGGAGAAACACACACAGATCCATCGGCCCCCGTGA 1131  
 QY 361 SerAlaLysGluTrpArgGlnArgSerLeuLeuGluLysGluLeuLeuPheAlaTrp 380  
 DB 1132 AGTGCACAGGAGTACCGACAGGAGCTACTTGGAAAAGAACTGCTTTCTTCTGCTTAT 1191  
 QY 381 AspValPheGlyLeuProPheValAspProAspSerTrpThrProGluGluValIlePro 400  
 DB 1192 GATGTTTTTGAATTCCTTTGTGATCCGGAATTCGAGCTCCAGAAAGAGTGAATCCC 1251  
 QY 401 LysArgLeuGlnGluLysGlnLysSerGluArgGluThrAlaValArgIleSerGlnGlu 420  
 DB 1252 AAGATTTGCAAGAGAAACAGAGTCCAGACGGAACACAGCCGTACGCACTCCCGAGAG 1311  
 QY 421 IleGlyAsnLeuMetLysGluIleGluThrLeuValGluGluLysThrLysGluSerLeu 440  
 DB 1312 ATTTGGAACTTATGAGAAATCGAGACCTTGTGGAAGAGAAGACCAAGAGTCACTG 1371  
 QY 441 AspValSerArgLeuThrArgGluGlyClyProLeuLeuTrpGluGlyIleSerLeuThr 460  
 DB 1372 GATGTGACGACACTGACCCGGGAAGGTGGCCCTGCTGTATGAAGGATCAGTCTCACC 1431  
 QY 461 MetAsnSerLysLeuLeuAsnGlyTrpGlnArgValValMetAspGlyValIleSerAsp 480  
 DB 1432 ATGAACTCCAAACTCTCTGAAATGGTATCCAGCGGTGGTGTGGACGCGTAACTCTGAC 1491  
 QY 481 HisGluCysGlnGluLeuGlnArgLeuThrAsnValAlaAlaThrSerGlyAspGlyTrp 500  
 DB 1492 CACGAGTGTGAGGAGTGCAGAGACTGACCAATGTGCAGCAACCTCAGGAGATGGGTAC 1551  
 QY 501 ArgGlyGlnThrSerProHisThrProAsnGluLysPheTrpGlyValThrValPheLys 520  
 DB 1552 CGGGTTCAGACTCCCACTACTCCCAATGAAAAGTTCTATGGTGTCTGCTTCTCAA 1611  
 QY 521 AlaLeuLysLeuGlyGlnGluLysValProLeuGlnSerAlaHisLeuTrpTrpAsn 540  
 DB 1612 GCGCTCAAGCTGGGGCAAGAGCAAGTTCTCTGCAAGTGGCCACCTGTACTACAAC 1671  
 QY 541 ValThrGluLysValArgArgIleMetGluSerTrpPheArgLeuAspThrProLeuTrp 560  
 DB 1672 GTGACGGAAGAAGTGGCGGCGCATCATGGAGTCTTACTTCCGCTGGATACGCCCTCTAC 1731  
 QY 561 PheSerTrpSerHisLeuValCysArgThrAlaIleGluGluValGlnAlaGluArgLys 580  
 DB 1732 TTTTCTTCTACTCTCATCTGGTGTGGCGCACTGCCATCGAAGAGGTCCAGGAGAGGAAG 1791  
 QY 581 AspAspSerHisProValHisValAspAsnCysIleLeuAsnAlaGluThrLeuValCys 600  
 DB 1792 GATGATGTCATCAGTCCAGTCCAGTCCAGTCCAGTCCAGTCCAGTCCAGTCCAGTCCAG 1851  
 QY 601 ValLysGluProProAlaTrpThrPheArgAspTrpSerAlaIleLeuTrpLeuAsnGly 620  
 DB 1852 GTCAAG 1911  
 QY 621 AspPheAspGlyGlyAsnPheTrpPheThrGluLeuAspAlaLysThrValThrAlaGlu 640

DB 1912 GACTTCGATCGCGAAACTTTTATTTCACTGAACCTGATGCCAAGACCGTGAAGCGCAGAG 1971  
QY 641 ValGlnProGlnCysGlyArgAlaValGlyPheSerSerGlyThrGluAsnProHisGly 660  
DB 1972 GTGACGCTCAGTGTGGAAGAGCGGTGGGATTCTTTCAGGCACTGAAACCCACATGGA 2031  
QY 661 ValIleAlaValThrArgGlyGlnArgCysAlaIleAlaLeuTrpPheThrLeuAspPro 680  
DB 2032 GTGAAGCTGTCCACAGGGGGCAGCGCTGTGCCATCGCCTGTGGTTACCTGGACCT 2091  
QY 681 ArgHisSerGluArgAspArgValGlnAlaAspAspLeuValIleValMetLeuPheSerPro 700  
DB 2092 CGACACAGCGCGGACAGCGGTGCAGGCGAGATGACCTGTGAGATGCTTTCAGCCCA 2151  
QY 701 GluGluMetAspLeuSerGlnGluGlnProLeuAspAlaGlnGlnGlyProProGluPro 720  
DB 2152 GAGAGATGGACCTCTCCAGGAGCAGCCCTGGATGCCAGCAGGGCCCCCGACCT 2211  
QY 721 AlaGlnGluSerLeuSerGlySerGluSerLysProLysAspGluLeu 736  
DB 2212 GCACAAGAGTCTCTCAGGACGTGAATCGAAGCCCAAGATGAGCTA 2259

## RESULT 2

ABZ11346  
ID ABZ11346 standard; cDNA; 2242 BP.  
XX  
AC ABZ11346;  
XX  
DT 20-JAN-2003 (first entry)  
XX  
DE Human polynucleotide SEQ ID NO 228.  
XX  
KW Human; genome mapping; gene therapy; food supplement; virus; fungus;  
KW cell-proliferative disorder; neurodegenerative disease; bacterial;  
KW Parkinson's disease; Alzheimer's disease; autoimmune disease;  
KW multiple sclerosis; diabetes; genetic disorder; wound; burn; infection;  
KW arthritis; cytostatic; immunomodulator; nootropic; neuroprotective;  
KW antiparkinsonian; antidiabetic; immunosuppressive; dermatological;  
KW haemostatic; vulnary; fungicide; antibacterial; virucide; protozoacide;  
KW antiarthritic; gene; ss.  
XX  
OS Homo sapiens.  
XX  
PN WO200270539-A2.  
XX  
PD 12-SEP-2002.  
XX  
PF 05-MAR-2002; 2002WO-US005095.  
XX  
PR 05-MAR-2001; 2001US-00799451.  
XX  
PA (HYSE-) HYSQ INC.  
XX  
PI Tang YT, Zhou P, Goodrich RW, Asundi V, Zhang J, Zhao QA, Ren F;  
PI Xue AJ, Yang Y, Ma Y, Yamazaki V, Chen R, Wang Z, Ghosh M;  
PI Wehrman T, Wang J, Wang D, Drmanac RT;  
XX  
DR WPI; 2002-759812/82.  
XX  
DR P-FSDS; ABP69129.  
XX  
PT New polynucleotides comprising sequences assembled from expressed  
PT sequence tags (ESTs), useful for treating cell-proliferative,  
PT neurodegenerative, autoimmune, genetic, myeloid or lymphoid, or platelet  
PT or coagulation disorders.  
XX  
PS Claim 1; SEQ ID NO 228; 1012pp + Sequence Listing; English.  
XX  
CC The invention relates to an isolated polynucleotide (I) comprising a  
CC nucleotide sequence selected from any of 948 sequences (ABZ11119-  
CC ABZ12066) or their mature protein coding portion, active domain coding  
CC protein or complementary sequences. The polynucleotides are useful for  
CC identifying expressed genes or for physical mapping of human genome. The

CC encoded polypeptides (ABP6902-ABP69849) are useful as molecular weight  
CC markers, as a food supplement, for generating antibodies, in medical  
CC imaging, screening and diagnostic assays and for treating cell-  
CC proliferative disorders (cancer), neurodegenerative diseases (Parkinson's  
CC or Alzheimer's disease), autoimmune diseases (multiple sclerosis,  
CC diabetes, lupus) genetic disorders, myeloid or lymphoid disorders,  
CC platelet or coagulation disorders, wound, burns, incision, ulcers, liver  
CC or lung fibrosis, infections (bacterial, viral, fungal, parasitic),  
CC arthritis, etc. Note: The sequence data for this patent did not form part  
CC of the printed specification, but was obtained in electronic format  
CC directly from WIPO at ftp.wipo.int/pub/published\_pct\_sequences

XX  
SQ Sequence 2242 BP; 518 A; 655 C; 625 G; 444 T; 0 U; 0 Other;

## Alignment Scores:

Pred. No.: 0 Length: 2242  
Score: 3861.00 Matches: 734  
Percent Similarity: 99.86% Conservative: 1  
Best Local Similarity: 99.73% Mismatches: 1  
Query Match: 99.77% Indels: 0  
DB: 6 Gaps: 0

US-10-045-815-4 (1-736) x ABZ11346 (1-2242)

QY 1 MetAlaValArgAlaLeuLysLeuLeuThrThrLeuLeuAlaValAlaAlaSer 20  
DB 32 ATGGCGGTACGCGGCTTGAAGCTGCTGACCACACTGCTGGCTGCTGGCGCTCCCTCC 91  
QY 21 GlnAlaGluValGluSerGluAlaGlyTrpGlyMetValThrProAspLeuLeuPheAla 40  
DB 92 CAAGCCGAGGTGCGATCCGAGGAGGATGGGCGCTGATGCGCTGATCTCTCTCC 151  
QY 41 GluGlyThrAlaAlaTyrAlaArgGlyAspTrpProGlyValValLeuSerMetGluArg 60  
DB 152 GAGGGGACCGCAGCTACTCGCGGGGACTGGCGCGGGTGGTCTCTGAGCATGGAACGG 211  
QY 61 AlaLeuArgSerArgAlaAlaLeuArgAlaLeuArgLeuArgCysArgThrGlnCysAla 80  
DB 212 GGGCTGCGCTCCGGGCGAGCCCTTCGCGCCCTTGGCTGGCTGGCGACCCAGTGTGCC 271  
QY 81 AlaAspPheProTrpGluLeuAspProAspTrpSerProSerProAlaGlnAlaSerGly 100  
DB 272 GCGGACTTCCGCTGGGAGCTGGACCCCGACTGGTCCCGACGCGCGCCGCGCTCGGGC 331  
QY 101 AlaGlyAlaLeuArgAspLeuSerPhePheGlyGlyLeuLeuArgAlaAlaCysLeu 120  
DB 332 GCGCGCGCGCTGCGCGACTGAGCTTCTCGGGGGCTTCTGCGTGGCGCTGCGCTGCGCTG 391  
QY 121 ArgArgCysLeuGlyProProAlaAlaHisSerLeuSerGluGluMetGluLeuPhe 140  
DB 392 CGCGCTGCTCGGGCGCGCGCGCCCTCGCTCAGCGAAGAGATGGAGTTC 451  
QY 141 ArgLysArgSerProTyrAsnTyrLeuGlnValAlaTyrPheLysIleAsnLysLeuGlu 160  
DB 452 CGCAAGCGGAGCCCTACACACTACTGCGAGTCCCTACTTCAAGATCAACAAGTTGGAG 511  
QY 161 LysAlaValAlaAlaAlaHisThrPhePheValGlyAsnProGluHisMetGluMetGln 180  
DB 512 AAAGCTGTGCTGCGAGCACACACCTTCTTCGTGGGCAATCTCTGAGCACATGGAATGAG 571  
QY 181 GlnAsnLeuAspTyrTyrGlnThrMetSerGlyValLysGluAlaAspPheLysAspLeu 200  
DB 572 CAGAACCTTAGACTATTACCAACCATGCTGGAGTGAAGAGAGGCGCGACTTCAAGGATCTT 631  
QY 201 GluThrGlnProHisMetGlnGluPheArgLeuGlyValArgLeuTyrSerGluGluGln 220  
DB 632 GAGACTCAACCCCATATGCAAGAAATTTGACTGGAGTGGCAGTCTACTCAGAGAACAG 691  
QY 221 ProGlnGluAlaValProHisLeuGluAlaAlaLeuGlnGluTyrPheValAlaTyrGlu 240  
DB 692 CCACAGGAAGCTGTGCCCCACCTAGAGGCGCGCTGCAAGAATACTTTGTGCCCTATGAG 751  
QY 241 GluCysArgAlaLeuCysGluGlyProTyrAspTyrAspGlyTyrAsnTyrLeuGluTyr 260

```

Db 752 GAGTGGCGTGCCTCTCGAAGGCGCCCTATGACTACGATGCGTACACTACCTTGAGTAC 811
Qy 261 AsnAlaAspLeuPheGlnAlaIleThrAspHisTyrIleGlnValLeuAsnCysLysGln 280
Db 812 AACGCTGACCTCTTCAGGCGCATACAGATCATATCATCCAGGTCTCACTGTAGGAG 871
Qy 281 AsnCysValThrGluLeuAlaSerHisProSerArgGluLysProPheGluAspPheLeu 300
Db 872 AACTGTGTACGGAGCTGTCTCCACCCCAAGTCAGAGAGGCCCTTTGAAGACTTCCCTC 931
Qy 301 ProSerHisTyrAsnTyrLeuGlnPheAlaTyrTyrAsnIleGlyAsnTyrThrGlnAla 320
Db 932 CCATCGCATATATAATTATCTGCGAGTTGCTTACTATTAACATTGGGAATATATACAGCT 991
Qy 321 GlyGluCysAlaLysThrTyrLeuLeuPhePheProAsnAspGluValMetAsnGlnAsn 340
Db 992 GGTGAATGTGCCAAGACTATCTTCTCTCTTCCCAATCAGCAGGTGATGAACCAAAAT 1051
Qy 341 LeuAlaTyrTyrAlaAlaMetLeuGlyGluGluHisThrArgSerIleGlyProArgGlu 360
Db 1052 TTGCGCTTATTATGAGCTATGCTTGGAGAGAAACACACAGATCCATCGGCCCCCGTGAG 1111
Qy 361 SerAlaLysGluTyrArgGlnArgSerLeuLeuGluLysGluLeuLeuPhePheAlaTyr 380
Db 1112 AGTCCCAAGGATACCGACAGCGAGCTACTGGAAAAGAACTGCTTTTCTTCGCTTAT 1171
Qy 381 AspValPheGlyIleProPheValAspProAspSerTrpThrProGluGluValIlePro 400
Db 1172 GATGTTTTTGGAAATTCGCTTTGTGGATCCGATTCATGGACTCCAGAGAAGTGAATCCC 1231
Qy 401 LysArgLeuGlnGluLysGlnLysSerGluArgGluThrAlaValArgIleSerGlnGlu 420
Db 1232 AAGAGATTGCNAGAAACACAGAGTCAAGACGGGAAACACCGGTACGCATCTCCAGGAG 1291
Qy 421 IleGlyAsnLeuMetLysGluIleGluThrLeuValGluGluLysThrLysGluSerLeu 440
Db 1292 ATTGGGAACCTTATGAAGAAATCGAGACCTTGTGGAGAGAAAGCAACCAAGGAGTCACTG 1351
Qy 441 AspValSerArgLeuThrArgGluGlyGlyProLeuLeuTyrGluGlyIleSerLeuThr 460
Db 1352 GATGTGACGACACTGACCCGGAAGGTGGCCCCCTGCTGTATGAAGCATCATCTCACC 1411
Qy 461 MetAsnSerLysLeuLeuAsnGlyTyrGlnArgValValMetAspGlyValIleSerAsp 480
Db 1412 ATGAATCCAAACTCTCTGAATGTTACCAGCGGTGTGTATGACGCGGTAACTCTGTAC 1471
Qy 481 HisGluCysGlnGluLeuGlnArgLeuThrAsnValAlaIleThrSerGlyAspGlyTyr 500
Db 1472 CACAGAGTGTACGAGAGCTGCAGAGACTACCAATGTGGCAGCAACCTCAGGAGATGCTAC 1531
Qy 501 ArgGlyGlnThrSerProHisThrProAsnGluLysPheTyrGlyValThrValPheLys 520
Db 1532 CGGGGTGAGACCTCCACACATCTCCCAATGAAGAAGTTCTATGGTGTCTCTTCAA 1591
Qy 521 AlaLeuLysLeuGlyGlnGluGlyLysValProLeuGlnSerAlaHisLeuTyrTyrAsn 540
Db 1592 GCCCTCAAGCTGGGGCAAGAGGCAAAAGTTCTCTGCAGAGTGCCCCACCTGTACTACAAC 1651
Qy 541 ValThrGluLysValArgArgGlyIleMetGluSerTyrPheArgLeuAspThrProLeuTyr 560
Db 1652 GTGACGAGAAAGTGGCGGCATCATGAGTCTTCTTCCGCTGGATACGCCCTCTTAC 1711
Qy 561 PheSerTyrSerHisLeuValCysArgThrAlaIleGluGluValGlnAlaGluArgLys 580
Db 1712 TTTTCTTACTCTCATCTGTGTGGTGGCGCACTGCTTCTTCCGCTGGATACGCCCTCTTAC 1771
Qy 581 AspAspSerHisProValHisValAspAsnCysIleLeuAsnAlaGluThrLeuValCys 600
Db 1772 GATGATAGTATCATCCAGTCCAGTGGACACTGCATCTCTGAATGCCAGAGACCTTGTGTGT 1831
Qy 601 ValLysGluProAlaTyrThrPheArgAspTyrSerAlaIleLeuTyrLeuAsnGly 620

```

```

Db 1832 GTCAAAGAGCCCCCAGCCCTACACCTTCCGCGACTACAGCGCATCTTTTACTTAATGGG 1891
Qy 621 AspPheAspGlyGlyAsnPheTyrPheThrGluLeuAspAlaLysThrValThrAlaGlu 640
Db 1892 GACTTCGATGGCGAAACTTTTATTTCACTGAATGGATGCCAAGACCGTGACGCGCAGAG 1951
Qy 641 ValGlnProGlnCysGlyArgAlaValGlyPheSerSerGlyThrGluAsnProHisGly 660
Db 1952 GTGCGACCTCATGTGTGGAAGAGCCGTGGGATTTCTTTCAGGACTGAAACCCCATGGA 2011
Qy 661 ValLysAlaValThrArgGlyGlnArgCysAlaIleAlaLeuTyrPheThrLeuAspPro 680
Db 2012 GTGAAGGCTGTCAACGAGGGGACGCGTGTGCCATCGCCTGTGTTTCCACCTTGACCT 2071
Qy 681 ArgHisSerGluArgAspArgValGlnAlaAspAspLeuValLysMetLeuPheSerPro 700
Db 2072 GCACACAGCAGCGGACAGGGGTGCAGGCAGATGACCTGGTGAAGATGCTCTTTCAGCCCA 2131
Qy 701 GluGluMetAspLeuSerGlnGluGlnProLeuAspAlaGlnGlnGlyProProGluPro 720
Db 2132 GAAGAGATGGACCTCTCCAGAGCAGCCCTGATGCCAGCAGGGCCCCCGGACCT 2191
Qy 721 AlaGlnGluSerLeuSerGlySerGlySerLysProLysAspGluLeu 736
Db 2192 GCACAAGAGTCTCTCTCAGGAGTGAATCGAAGCCCAAGGATGAGCTA 2239

```

RESULT 3  
AAS17573  
ID AAS17573 standard; cDNA; 2211 BP.  
XX  
AC AAS17573;  
XX  
DT 26-FEB-2002 (first entry)  
XX  
DE DNA encoding novel secreted protein #2.  
XX  
KW Secreted protein; cytostatic; immunosuppressive; vulnary; vaccine;  
KW antinflammatory; neuroprotective; nephrotropic; cardiovascular; human;  
KW cancer; autoimmune disease; wound healing disorder; infection;  
KW haematopoietic disorder; inflammatory disorder; infertility;  
KW neurological disease; psychiatric disease; cardiovascular disease;  
KW respiratory disease; renal; gastrointestinal; ss.  
XX  
OS Homo sapiens.  
XX  
FH Key Location/Qualifiers  
CDS 1..2211  
FT /\*tag= a  
FT /product= "Human secreted protein"  
XX  
PN WO200179454-A1.  
XX  
PD 25-OCT-2001.  
XX  
PP 11-APR-2001; 2001WO-US011797.  
XX  
PR 13-APR-2000; 2000US-0196603P.  
PR 24-APR-2000; 2000US-0199417P.  
XX  
PA (SMIK ) SMITHKLINE BEECHAM CORP.  
PA (SMIK ) SMITHKLINE BEECHAM PLC.  
PI Agarwal P, Murdoch PR, Rizvi SK, Smith RF, Xiang Z;  
XX  
DR WPI; 2002-061975/08.  
DR P-PSDB; AAU09861.  
XX  
PT New secreted proteins or polypeptides, useful for treating e.g. cancer,  
PT autoimmune diseases, wound healing disorder, infections, hematopoietic  
PT disorders, inflammatory disorders, infertility, cancer.  
XX  
PS Claim 2; Page 35-36; 92pp; English.  
XX

CC The invention relates to an isolated novel secreted polypeptide (I) and  
CC polynucleotide (II). (I) and (II) are useful for treating cancer,  
CC autoimmune diseases, wound healing disorder, infections, haematopoietic  
CC disorders, inflammatory disorders, infertility, neurological and  
CC psychiatric diseases, cardiovascular diseases, respiratory diseases,  
CC renal diseases, or gastrointestinal diseases. These may also be used to  
CC treat diseases, abnormalities and disorders caused by abnormal  
CC expression, production, function and/or metabolism of the genes, as  
CC vaccines for inducing immunological response in a mammal, and in  
CC screening methods for detecting the effect of added compounds on the  
CC production of mRNA and polypeptide in cells. The polypeptides can be used  
CC as immunogens to produce antibodies immunospecific for the polypeptides,  
CC and to identify membrane-bound or soluble receptors. The polynucleotides  
CC may be used as diagnostic reagents, in chromosome localisation studies,  
CC and in tissue expression studies. The present sequence represents the  
CC coding sequence of novel human secreted protein #2  
XX  
SQ Sequence 2211 BP; 512 A; 649 C; 612 G; 438 T; 0 U; 0 Other;

Alignment Scores:  
Pred. No.: 0 Length: 2211  
Score: 3846.00 Matches: 733  
Percent Similarity: 99.59% Conservative: 0  
Best Local Similarity: 99.59% Mismatches: 3  
Query Match: 99.38% Indels: 0  
DB: 6 Gaps: 0

US-10-045-815-4 (1-736) x AAS17573 (1-2211)

QY	1	MetAlaValArgAlaLeuLysLeuThrThrLeuLeuAlaValValAlaAlaLaser	20
DB	1	ATGGCGGTACGCGGTTGAAGCTGTGACACACTGCTGGCTGTGCGCCGTGCTCC	60
QY	21	GlnAlaGluValGluSerGluAlaGlyTrpGlyMetValThrProAspLeuLeuPheAla	40
DB	61	CAAGCCGAGTCCGAGTCGAGGCGAGGATGGGGCATGGTGACGCTGATCTGCTCTCGCC	120
QY	41	GluGlyThrAlaAlaTyrAlaArgGlyAspTrpProGlyValValLeuSerMetGluArg	60
DB	121	GAGGGACCGACCTACCGCGCGGGGACTGGCCCGGGGTGCTCGTGAATGGAACGG	180
QY	61	AlaLeuArgSerArgAlaAlaLeuAlaLeuArgLeuArgCysArgThrGlnCysAla	80
DB	181	GCCTCGGCTCCGGGCGAGCCCTCCGGCCCTTCGGCTCGGCTCGGACCCAGTGTGCC	240
QY	81	AlaAspPheProTrpGluLeuAspProAspTrpSerProSerProAlaGlnAlaSerGly	100
DB	241	GCCGACTTCCGTTGGGAGCTGACCCCGACTGGTCCCGCCAGCCCGCCAGGCTCGGGC	300
QY	101	AlaGlyAlaLeuArgAspLeuSerPheGlyGlyLeuLeuArgAlaAlaCysLeu	120
DB	301	GCCCGCCCTTCGCGACCTGAGCTTCTTCGGGGGCTTCTGGGTGCGCTGCTGCTGCTG	360
QY	121	ArgArgCysLeuGlyProProAlaAlaHisSerLeuSerGluGluMetGluLeuGluPhe	140
DB	361	CGCCGCTGCTCGGCGCGCGCGCCGCTCGCTCAGCGAGAGATGGAGTTC	420
QY	141	ArgLysArgSerProTyrAsnTyrLeuGlnValAlaTyrPheLysIleAsnLysLeuGlu	160
DB	421	CGCAAGCGGAGCCCTACCACTACCTGCGAGTGGCTTCTTCAAGATCAACAAGTTGGAG	480
QY	161	LysAlaValAlaAlaAlaHisThrPhePheValGlyAsnProGluHisMetGluMetGln	180
DB	481	AAAGCTGTGCTGCAGACACACACTTCTCGTGGGCAATCCTGAGCACATGGAAATGCAG	540
QY	181	GlnAsnLeuAspTyrTyrGlnThrMetSerGlyValLysGluAlaAspPheLysAspLeu	200
DB	541	CAGAACCCTAGACTATTACCAACCATGTCTGGAGTGAAGGAGGCCGACTTCAAGGATCTT	600
QY	201	GluThrGlnProHisMetGlnGluPheArgLeuGlyValValArgLeuTyrSerGluGln	220
DB	601	GAGACTCAACCCCATATGCAAGAAATTCGACTGGGAGTGGGACTCTACTACAGAGNACAG	660

QY	221	ProGlnGluAlaValProHisLeuGluAlaAlaLeuGlnGluTyrPheValAlaTyrGlu	240
DB	661	CCACAGGAGGTGTGTGCCCACTAGAGCGGGCTGCAAGAAATACTTTGTGGCTATGAG	720
QY	241	GluCysArgAlaLeuCysGluGlyProTyrAspTyrAspGlyTyrAsnTyrLeuGluTyr	260
DB	721	GAGTCCGCTGCTGTGGAAGGCGCTATGACTACGATGGCTACAACTACCTTGAGTAC	780
QY	261	AsnAlaAspLeuPheGlnAlaIleThrAspHisTyrIleGlnValLeuAsnCysLysGln	280
DB	781	AACGCTGACCTTCTCCAGGCCATCACAGATCATATCCAGGCTCTCACTGTAAAGAG	840
QY	281	AsnCysValThrGluLeuAlaSerHisProSerArgGluLysProPheGluAspPheLeu	300
DB	841	AACGTGTACGGAGCTTGTCTCCACCACCAAGTCGAGAGAAGCCCTTTGAAGACTTCCTC	900
QY	301	ProSerHisTyrAsnTyrLeuGlnPheAlaTyrTyrAsnIleGlyAsnTyrThrGlnAla	320
DB	901	CCATCGCATTAATAATATCTGCAGTTTGCCTACTATACATGGGAAATATACACAGCT	960
QY	321	GlyGluCysAlaLysThrTyrLeuLeuPhePheProAsnAspGluValMetAsnGlnAsn	340
DB	961	GTTGAATGTGCCAAGACCTATCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT	1020
QY	341	LeuAlaTyrTyrAlaAlaMetLeuGlyGluHisThrArgSerIleGlyProArgGlu	360
DB	1021	TTGGCTATTATGACGCTATGCTTGGAGAAGACACACAGATCCATCGGCCCCCGTGAG	1080
QY	361	SerAlaLysGluTyrArgGlnArgSerLeuLeuGluLysGluLeuPhePheAlaTyr	380
DB	1081	AGTGCAAGGAGTACCGACGAGGCTACTTGGAAAAGAACTGCTTTCTTCTGCTTAT	1140
QY	381	AspValPheGlyIleProPheValAspProAspSerTrpThrProGluGluValIlePro	400
DB	1141	GATGTTTTTGGAAATTCCTTTGTGATCCGATTCATGGATCCAGATCCAGAGAAGTATTCC	1200
QY	401	LysArgLeuGlnGluLysGlnLysSerGluArgGluThrAlaValArgIleSerGlnGlu	420
DB	1201	AAGAGATTGCAAGAAACACAAAGTCAAGACGGGAAACAGCCGACCGATCTCCAGAGG	1260
QY	421	IleGlyAsnLeuMetLysGluIleGluThrLeuValGluGluLysThrLysGluSerLeu	440
DB	1261	ATTGGAACTTATGAAGAAATCGAGACCTTGTGGAAGAGAAGACCAAGAGTCACTG	1320
QY	441	AspValSerArgLeuThrArgGluGlyGlyProLeuLeuTyrGluGlyIleSerLeuThr	460
DB	1321	GATGTGACAGACTGACCCGGAAAGTGGCCCTCTGCTGTATGAAGCATCAGTCTCACC	1380
QY	461	MetAsnSerLysLeuLeuAsnGlyTyrGlnArgValValMetAspGlyValIleSerAsp	480
DB	1381	ATGAACTCCAACTCTGATGGTTCCAGCGGGTGTGATGGACGGGTAATCTCTGAC	1440
QY	481	HisGluCysGlnGluLeuGlnArgLeuThrAsnValAlaAlaThrSerGlyAspGlyTyr	500
DB	1441	CACGAGTGTGAGGAGCTGCAGAGACTGACCAATGTGCGAGCAACCTCAGGAGATGGCTAC	1500
QY	501	ArgGlyGlnThrSerProHisThrProAsnGluLysPheTyrGlyValThrValPheLys	520
DB	1501	CGGGTTCAGACTCCCTCCACATCTCCCAATGAAAGTTCTATGGTGTCTCTCTCTCTCT	1560
QY	521	AlaLeuLysLeuGlyGlnGluLysValProLeuGlnSerAlaHisLeuTyrTyrAsn	540
DB	1561	GCCCTCAAGCTGGGCGAAGAGGCAAAAGTTCTCTGCGAGAGTGCCCACTCTACTACAAC	1620
QY	541	ValThrGluLysValArgArgIleMetGluSerTyrPheArgLeuAspThrProLeuTyr	560
DB	1621	GTGACGGAGAAGGTCGCGCGCATCATGGAGTCTTCTCTGCGCTGGATACGCCCTCTAC	1680
QY	561	PheSerTyrSerHisLeuValCysArgThrAlaIleGluGluValGlnAlaGluArgLys	580
DB	1681	TTTTCTTCT	1740
QY	581	AspAspSerHisProValHisValAspAsnCysIleLeuAsnAlaGluThrLeuValCys	600



Db 1741 GATGATAGTTCATCCAGTCCAGTGGACACTGCTGATCTGTAATCCGAGACCCCTCGTGTGT 1800  
 Qy 601 VallysGluProAlaTyrThrPheArgAspTyrSerAlaIleLeuTyrLeuAsnGly 620  
 Db 1801 GTCAAGAGAGCCCGCAGCTACACCTCCGCGACTACAGCCCATCTTTACCTAAATGGG 1860  
 Qy 621 AspPheAspGlyAenPheTyrPheThrGluLeuAspAlaIleTyrValThrAlaGlu 640  
 Db 1861 GACTTCGATCGCGAATCTTTATTTTCACTGAATCGGATCCGAGACCGGTGCGGAGAG 1920  
 Qy 641 ValGlnProGlnCysGlyArgAlaValGlyPheSerSerGlyThrGluAsnProHisGly 660  
 Db 1921 GTGCAGCTCAGTGTGAAGAGCGGTGGATTCTCTCAGGCACTGAAACCCACATGGA 1980  
 Qy 661 VallysAlaValThrArgGlyGlnArgCysAlaIleAlaLeuTyrPheThrLeuAspPro 680  
 Db 1981 GTGAAGGCTGTCAACAGGGGCGAGCGCTGTGCCATCGCCCTGTGTGTTCACTCGGACCT 2040  
 Qy 681 ArgHisSerGluArgAspArgValGlnAlaAspLeuVallysMetLeuPheSerPro 700  
 Db 2041 CGACACAGCAGCGGAGCGGTGCGAGATGACCTGTGTAAGATGCTCTTCACGCCCA 2100  
 Qy 701 GluGluMetAspLeuSerGlnGlnProLeuAspAlaGlnGlnGlyProProGluPro 720  
 Db 2101 GAAGAGATGACCTCTCCAGGAGCAGCCCTCGATGCCAGCAGGCTCCCCCGAACCT 2160  
 Qy 721 AlaGlnGluSerLeuSerGlySerGluSerLysProLysAspGluLeu 736  
 Db 2161 GCACAAGATCTCTCTCAGGAGTGAATCGAAGCCCAAGATGAGCTA 2208

RESULT 4  
 AAF93800  
 ID AAF93800 standard; cDNA; 2563 BP.

AC AAF93800;  
 XX  
 DT 23-MAY-2001 (first entry)  
 XX  
 DE Human cDNA encoding a membrane or secretory protein clone PSEC0109.  
 XX  
 KW Human; secretory protein; membrane protein; vaccine; gene therapy;  
 XX  
 OS rheumatoid arthritis; diabetes; ss.  
 XX  
 OS Homo sapiens.  
 XX  
 FN EPI067182-A2.  
 XX  
 PD 10-JAN-2001.  
 XX  
 PF 07-JUL-2000; 2000EP-00114090.  
 XX  
 PR 08-JUL-1999; 99JP-00194179.  
 PR 11-JAN-2000; 2000JP-00118775.  
 PR 02-MAY-2000; 2000JP-00183766.  
 XX  
 PA (HELI-) HELIX RES INST.  
 XX  
 PI Ota T, Isogai T, Nishikawa T, Kawai Y, Sugiyama T, Hayashi K;  
 XX  
 DR WPI; 2001-093989/11.  
 DR P-PSDB; AAB88373.  
 XX

PT Nucleic acids encoding secretory proteins/membrane proteins, useful in  
 PT gene therapy or as candidate target molecules in drug development.  
 XX  
 PS Claim 1; SEQ ID NO 113; 609pp + Sequence Listing; English.

CC This invention relates to nucleic acid sequences AAF93744 - AAF93916  
 CC which encode human secretory or membrane proteins represented by AAB88317  
 CC - AAB88419. Included in the invention are primers AAF93917 - AAF94295 and  
 CC AAF52232 - AAF62235 which are used to isolate the cDNA sequences of the  
 CC invention. The invention also includes methods for the production of

CC antibodies directed against the proteins, and cDNA sequences, which can  
 CC be used in vaccines. The polynucleotide sequences can be used in gene  
 CC therapy. The polynucleotide sequences and the proteins they encode may be  
 CC used in the prevention, treatment and diagnosis of diseases associated  
 CC with inappropriate secretory protein/membrane protein expression. The  
 CC nucleic acids and complementary sequences may also be used as DNA probes  
 CC in diagnostic assays (e.g. polymerase chain reactions (PCR)) to detect  
 CC and quantitate the presence of similar nucleic acid sequences in samples.  
 CC They may also be used to study the expression and function of secretory  
 CC proteins/membrane polypeptides and their role in metabolism. The  
 CC polypeptides may be used as antigens in the production of antibodies  
 CC against them and in assays to identify modulators (agonists and  
 CC antagonists) of expression and activity. The antibodies and antagonists  
 CC may also be used as therapeutic agents to down regulate expression and  
 CC activity. The antibodies may also be used as diagnostic agents for  
 CC detecting the presence of the polypeptides in samples (e.g. by enzyme  
 CC linked immunosorbent assay (ELISA). Examples of diseases which may be  
 CC treated include rheumatoid arthritis and diabetes  
 XX

SQ Sequence 2563 BP; 589 A; 753 C; 713 G; 508 T; 0 U; 0 Other;

#### Alignment Scores:

Pred. No.: 0 Length: 2563  
 Score: 3846.00 Matches: 733  
 Percent Similarity: 99.59% Conservative: 0  
 Best Local Similarity: 99.59% Mismatches: 3  
 Query Match: 99.38% Indels: 0  
 DB: 5 Gaps: 0

US-10-045-815-4 (1-736) x AAF93800 (1-2563)

Qy 1 MetAlaValArgAlaLeuLysLeuLeuThrThrLeuLeuAlaValValAlaAlaSer 20  
 Db 31 ATGGCGGTACGCGGCTTGAAGCTGCTGACCACACTGCTGGCTGTGTGGCGCTGCCCTCC 90  
 Qy 21 GlnAlaGluValGluSerGluAlaGlyTyrGlyMetValThrProAspLeuLeuPheAla 40  
 Db 91 CAAGCCGAGGTGCGAGTCCGAGGAGGATGGGGCATGTGTGAGCCCTGATCTGCTTCGCC 150  
 Qy 41 GluGlyThrAlaAlaTyrAlaArgGlyAspTyrProGlyValValLeuSerMetGluArg 60  
 Db 151 GAGGGGACCGCAGCCTACGCGCGCGGGACTGGCCCGGGGTGGTCTCTGAGCATGGAACGG 210  
 Qy 61 AlaLeuArgSerArgAlaAlaLeuArgAlaLeuArgGluLeuArgCysArgThrGlnCysAla 80  
 Db 211 GCGCTGCGCTCCCGGCGAGCCCTCCGCGCCCTTCGCTGCGCTCCGCGACCCAGTGTGCC 270  
 Qy 81 AlaAspPheProTyrGluLeuAspProAspTyrSerProSerProAlaGlnAlaSerGly 100  
 Db 271 GCCGACTTCCCGTGGGAGCTGACCCCGACTGTTCCCGCCAGCCCGCCGCGCTCCGGC 330  
 Qy 101 AlaGlyAlaLeuA:GAspLeuSerPhePheGlyGlyLeuLeuArgArgAlaAlaCysLeu 120  
 Db 331 GCGCGCGCCCTGCGCGACCTGAGCTTCTTCGGGGGCGCTTCTGCGCTCGCGCTGCTGCTG 390  
 Qy 121 ArgArgCysLeuGlyProProAlaAlaHisSerLeuSerGluGluMetGluLeuGluPhe 140  
 Db 391 GCCCGCTCCCTCGGGCGCGCGCCCGCCACTCTGCTCAGCGAGAGATGGAGCTGGAGTTC 450  
 Qy 141 ArgLysArgSerProTyrAsnTyrLeuGlnValAlaTyrPheLysIleAsnLysLeuGlu 160  
 Db 451 CGCAAGCGGAGCCCTACAACTACTCTGAGGTGCGCTTCTCAAGATCAACAAGATTGGAG 510  
 Qy 161 LysAlaValAlaAlaAlaHisThrPhePheValGlyValAsnProGluHisMetGluMetGln 180  
 Db 511 AAAGCTGTGTGTCAGGACACACCTTCTTCGTGGGCAATCTCTGAGCATGGAAATTCAG 570  
 Qy 181 GlnAsnLeuAspTyrTyrGlnThrMetSerGlyValLysGluAlaAspPheLysAspLeu 200  
 Db 571 CAGAACCTTAGACTATTACCAAAACCATGTCTGGAGTGAAGGAGGCGGACTTCAAGGATCTT 630  
 Qy 201 GluThrGlnProHisMetGlnGluPheArgLeuGlyValArgLeuTyrSerGluGln 220



631 GAGACTCAACCCCATATGCAAGAAATTTGACTGGGAGTGGGACTCTACTCAGAGGAACAG 690  
221 ProGlnGluAlaValProHisLeuGluAlaAlaLeuGlnGluTyrPheValAlaTyrGlu 240  
691 CCACAGAGAGCTGTGGCCCACTAGAGGGGGCTGCAAGAAATCTTTTGGCCCTATGAG 750  
241 GluCysArgAlaLeuGluGluGluGluGluGluGluGluGluGluGluGluGluGluGlu 260  
751 GAGTGGCGGCTCTGCGAGGGCCCTATGACTACGATGGCTACCACTACCTTGGATAC 810  
261 AsnAlaAspLeuPheGlnAlaIleThrAspHisTyrIleGlnValLeuAsnCysTyrGln 280  
811 AACGCTGACCTCTCCAGGCGCATCAGACATCATTACATCAGGCTCCTCAACTGTAAGCAG 870  
281 AsnCysValThrGluLeuAlaSerHisProSerArgGluLysProPheGluAspPheLeu 300  
871 AACTGTGTACGGAGCTGTCTTCCCAAGTCGAGAGAGCCCTTTGAAGACTTCCTC 930  
301 ProSerHisTyrAsnTyrLeuGlnPheAlaTyrTyrAsnIleGlyAsnTyrThrGlnAla 320  
931 CCATCGCATTAATAATTATCTGCAGTTTGCCTACTATAACATTGGGAATTATACAGGCT 990  
321 GlyGluCysAlaLysThrTyrLeuLeuPhePheProAsnAspGluValMetAsnGlnAsn 340  
991 GTTGAATGTGCCAAGACCTATCTTCTTCTTCCCAATGAGAGGTGATGAACCAAAAT 1050  
341 LeuAlaTyrTyrAlaAlaMetLeuGluGluGluHisThrArgSerIleGlyProArgGlu 360  
1051 TTGGCTTATATGCACTATGCTTGGAGAGAAACACACCATGATCCATCGCCCCCGTGAG 1110  
361 SerAlaLysGluTyrArgGlnArgSerLeuLeuGluLysGluLeuLeuPhePheAlaTyr 380  
1111 AGTGCCAGAGGATCCGACAGCGAAGCTACTGCAAAAGAACTGCTTTCTTCTGCTTAT 1170  
381 AspValPheGlyIleProPheValAspProAspSerTyrThrProGluGluValIlePro 400  
1171 GATGTTTTTGGAAATCCCTTTTGGATCCGGATTCATGGACTCCAGAAAGATGATTCCC 1230  
401 LysArgLeuGlnGluLysGlnLysSerGluArgGluThrAlaValArgIleSerGlnGlu 420  
1231 AAGAGATTGCAAGAGAAACAGAGTCCAGAACGGAAACAGCGTACGCTATCCCGAGGAG 1290  
421 IleGlyAsnLeuMetLysGluIleGluThrLeuValGluGluLysThrLysGluSerLeu 440  
1291 ATTGGGAACCTTATGAAGGAATTCGAGACCTTGTGGAAGAGAGAACCAAGAGATCACTG 1350  
441 AspValSerArgLeuThrArgGluGluGluGluGluGluGluGluGluGluGluGluGlu 460  
1351 GATGTGAGCAGACTGACCCGGGAAGTGGCCCCCTGCTGTATGAAGGCATCAGTCTCACC 1410  
461 MetAsnSerLysLeuLeuAsnGlyTyrGlnArgValValMetAspGlyValIleSerAsp 480  
1411 ATGAACCTCCAAACTCCTGAATGTTCCAGCGGGTGGTGTATGAGCGCGCTAATCTCTGAC 1470  
481 HisGluCysGlnGluLeuGlnArgLeuThrAsnValAlaAlaThrSerGlyAspGlyTyr 500  
1471 CAGGAGTGTACGAGGCTGTGAGAGCTGACCAATGTGGCAGCAACCTCAGGAGATGGCTAC 1530  
501 ArgGlyGlnThrSerProHisThrProAsnGluLysPheTyrGlyValThrValPheLys 520  
1531 CGGGGTACAGACTCCCAACATCTCCCAATGAAAGTTCTATGGTGTCTACTGCTTCAA 1590  
521 AlaLeuLysLeuGlnGluGluLysValProLeuGlnSerAlaHisLeuTyrTyrAsn 540  
1591 GCCCTCAAGCTGGGGGAAGAACAAAGTTCTCTCAGAGTGGCCACCTGTACTACAC 1650  
541 ValThrGluLysValArgGluMetGluSerTyrPheArgLeuAspThrProLeuTyr 560  
1651 GTGACGGAGAGTGGGGCGCATCAGGAGTCTTCTTCCGCTGGATACGCCCTCTTAC 1710  
561 PheSerTyrSerHisLeuValCysArgThrAlaIleGluGluValGlnAlaGluArgLys 580  
1711 TTTTCTCTACTCTCATCTGTGTGCCGACTGCCATCGAAGAGGTCCAGGAGAGGAG 1770

581 AspAspSerHisProValHisValAspAsnCysIleLeuAsnAlaGluThrLeuValCys 600  
1771 GATGATAGTCAATCCAGTCCACGTGGACAACTGCATCTCTGAATCCGAGACCTCGTGTGT 1830  
601 ValLysGluProProAlaTyrThrPheArgAspTyrSerAlaIleLeuTyrLeuAsnGly 620  
1831 GTCAAAGAGCCCGCAGCTACACCTTCGGGACTACAGGGCCATCTTTACCTAAATGGG 1890  
621 AspPheAspGlyGlyAsnPheTyrPheThrGluLeuAspAlaLysThrValThrAlaGlu 640  
1891 GACTTCGATGGCGAAACTTTTATTTTCACTGAATGGATGCAAGACCGTGCAGCGCAGAG 1950  
641 ValGlnProGlnCysGlyArgAlaValGlyPheSerSerGlyThrGluAsnProHisGly 660  
1951 GTGAGACCTCAGTGTGGAGAGCGGTGGGATCTCTTCAGSCACTGAAACCCACATGGA 2010  
661 ValLysAlaValThrArgGlyGlnArgCysAlaIleAlaLeuIlePheThrLeuAspPro 680  
2011 GTGAAGGCTGTCAACAGGGGCGAGCTGTGCCATCGCCCTGTGTTTCACTCGACCT 2070  
681 ArgHisSerGluArgAspArgValGlnAlaAspAspLeuValLysMetLeuPheSerPro 700  
2071 CGACACAGCAGCGGGACAGGGTGCAGCAGATCACTGGTGAAGATGCTCTTCAGGCCCA 2130  
701 GluGluMetAspLeuSerGlnGluGlnProLeuAspAlaGlnGlnGlyProProGluPro 720  
2131 GAAGAGATGAGCTCTCCAGAGAGACCCCTCGATGCCAGCAGGCGCCCCCGAACCT 2190  
721 AlaGlnGluSerLeuSerGlySerGluSerLysProLysAspGluLeu 736  
2191 GCACAAAGAGTCTCTCTCAGGCGAGTGAATCGAAGCCCAAGGATGAGCTA 2238

RESULT 5

AAH14589  
ID AAH14589 standard; cDNA; 2585 BP.  
XX  
AC AAH14589;  
XX  
DT 26-JUN-2001 (first entry)  
DE Human cDNA sequence SEQ ID NO:12193.  
XX  
KW Human; primer; detection; diagnosis; antisense therapy; gene therapy; ss.  
XX Homo sapiens.  
XX EP1074617-A2.  
XX  
PD 07-FEB-2001.  
XX  
PF 28-JUL-2000; 2000EP-00116126.  
XX  
PR 29-JUL-1999; 99JP-00248036.  
PR 27-AUG-1999; 99JP-00300253.  
PR 11-JAN-2000; 2000JP-00118776.  
PR 02-MAY-2000; 2000JP-00183767.  
PR 09-JUN-2000; 2000JP-00241899.  
XX  
PA (HELI-) HELIX RES INST.  
XX  
PI Ota T, Isogai T, Nishikawa T, Hayashi K, Saito K, Yamamoto J;  
PI Ishii S, Sugiyama T, Wakamatsu A, Negai K, Otsuki T;  
XX  
XX WPI; 2001-318749/34.

Primer sets for synthesizing polynucleotides, particularly the 5602 full-length cDNAs defined in the specification, and for the detection and/or diagnosis of the abnormality of the proteins encoded by the full-length cDNAs.

Claim 8; SEQ ID NO 12193; 2537pp + Sequence Listing; English.

CC The present invention describes primer sets for synthesizing 5602 full-length cDNAs defined in the specification. Where a primer set comprises: (a) an oligo-dT primer and an oligonucleotide complementary to the complementary strand of a polynucleotide which comprises one of the 5602 nucleotide sequences defined in the specification, where the oligonucleotide comprises at least 15 nucleotides; or (b) a combination of an oligonucleotide comprising a sequence complementary to the complementary strand of a polynucleotide which comprises a 5'-end sequence and an oligonucleotide comprising a sequence complementary to a polynucleotide which comprises a 3'-end sequence, where the oligonucleotide comprises at least 15 nucleotides and the combination of the 5'-end sequence/3'-end sequence is selected from those defined in the specification. The primer sets can be used in antisense therapy and in gene therapy. The primers are useful for synthesizing polynucleotides, particularly full-length cDNAs. The primers are also useful for the detection and/or diagnosis of the abnormality of the proteins encoded by cDNAs easily without any specialised methods. AAH03166 to AAH13628 and AAH13633 to AAH18742 represent human cDNA sequences; AAH92446 to AAH95893 represent human amino acid sequences; and AAH13629 to AAH13632 represent oligonucleotides, all of which are used in the exemplification of the present invention

XX Sequence 2585 BP; 590 A; 759 C; 723 G; 513 T; 0 U; 0 Other;

#### Alignment Scores:

Pred. No.: 0 Length: 2585  
Score: 3846.00 Matches: 733  
Percent Similarity: 99.59% Conservative: 0  
Best Local Similarity: 99.59% Mismatches: 3  
Query Match: 99.38% Indels: 0  
DB: 4 Gaps: 0

US-10-045-815-4 (1-736) x AAH14589 (1-2585)

QY	1	MetAlaValArgAlaLeuLysLeuLeuThrThreLeuAlaValValAlaAlaSer	20	593	CAGAACCTAGACTATTACCAAAACCATGTCTGGAGTGAAGGAGCGGACCTTCAAGGATCTT	652
DB	53	ATGGCGGTACCGCGGTGAAGCTGCTGACCACTGCTGCTGGTGGCGGTGCTCTCC	112	201	GluThrGlnProHisMetGlnGluPheArgLeuGlyValArgLeuTyrSerGluGluGln	220
QY	21	GlnAlaGluValGluSerGluAlaGlyTrpGlyMetValThrProAspLeuLeuPheAla	40	653	GAGATTCACCCCATATCAAGAAATTCGACTGGGAGTGGCGACTCTACTCAGAGGAACAG	712
DB	113	CAAGCCGAGGTCGAGTCCGAGGCGAGGATGGGCGATGGTGCCTGCTGCTCTCGCC	172	221	ProGlnGluAlaValProHisLeuGluAlaAlaLeuGlnGluTyrPheValAlaTyrGlu	240
QY	41	GluGlyThrAlaAlaTyrAlaArgGlyAspTrpProGlyValValLeuSerMetGluArg	60	713	CCACAGGAAGCTGTGCCCCACCTAGAGCGCGCTGCAAGAAATACCTTTGTGGCTATGAG	772
DB	173	GAGGGACCGCAGCTACCGCGCGGGGATGGCGCGGGTGTCTGAGCATGGAAACGG	232	241	GluCysArgAlaLeuCysGluGlyProTyrAspTyrAspGlyTyrAsnTyrLeuGluTyr	260
QY	61	AlaLeuArgSerArgAlaAlaLeuArgAlaLeuArgLeuArgCysArgThrGlnCysAla	80	773	GAGTGGCGTGGCTCTGCGAAGGGCCCTATGACTACGATGGCTCAACTACCTTTGAGTAC	832
DB	233	CGCGTGGCTCCCGGCGAGCCCTCCGCGCCCTGCGCTGCGCTGCCACCCAGTGTGCC	292	261	AsnAlaAspLeuPheGlnAlaIleThrAspHisTyrIleGlnValLeuAsnCysLysGln	280
QY	81	AlaAspPheProTrpGluLeuAspProAspTrpSerProSerProAlaGlnAlaSerGly	100	833	RACGCTGACCTTCCAGGCCCATCACAGATCATTTACATCCAGGTCCTCAACTGTAAAGCAG	892
DB	293	GCCGACTTCCCGTGGGAGCTGGACCCCGACTGGGTCCCGCCAGCCCGGCGCCAGGGCTCGGGC	352	281	AsnCysValThrGluLeuAlaSerHisProSerArgGluLysProPheGluAspPheLeu	300
QY	101	AlaGlyAlaLeuArgAspLeuSerPheGlyLeuLeuArgArgAlaAlaCysLeu	120	893	AACTGTGTCAAGGAGCTTGTCTCCACCAAGTCGAGAGAAGCCCTTTGAAGAGCTTCCTC	952
DB	353	GCCGCCCGCTCGCGACCTGAGCTTCTTGGGGGCTTCTGGGTGGCGTGGCTGCTGCTG	412	301	ProSerHisTyrAsnTyrLeuGlnPheAlaTyrIleGlnValLeuAsnTyrThrGlnAla	320
QY	121	ArgArgCysLeuGlyProProAlaAlaHisSerLeuSerGluGluMetGluLeuGluPhe	140	953	CCATCGCATTAATTAATTTCTGAGTTTGCCTACTATACATTTGGAAATATACACAGCT	1012
DB	413	CGCGCTGCTCGCGCGCGCGCGCCCTCGCTCGCTCGCTCGCTCGCTCGCTCGCTCGCTG	472	321	GlyGluCysAlaLysThrTyrLeuLeuPhePheProAsnAspGluValMetAsnGlnAsn	340
QY	141	ArgLysArgSerProTyrAsnTyrLeuGlnValAlaTyrPheLysIleAsnLysLeuGlu	160	1013	GTTGATGTGCCAAGACCTATCTCTCTCTCCCAATGACGAGGTGATGAACCAAAAT	1072
DB	473	CGCAAGCGGAGCCCTACAACCTACCTCGAGTGGCTTCTCAAGATCAACAAGTTGGAG	532	341	LeuAlaTyrTyrAlaAlaMetLeuGlyGluGluHisThrArgSerIleGlyProArgGlu	360
QY	161	LysAlaValAlaAlaHisThrPheValGlyAsnProGluHisMetGluMetGln	180	1073	TTGGCTATTATGACGATATCTTGGAGAGAACACACACAGATCCATCGGCCCCCGTGAG	1132
DB	533	AAAGCTGTGTGCGAGCACACACCTTCTTGTGGGCAATCCTGAGCACATGGAAATGACG	592	361	SerAlaLysGluTyrArgGlnArgSerLeuLeuGluLysGluLeuLeuPheAlaTyr	380
QY	181	GlnAsnLeuAspTyrTyrGlnThrMetSerGlyValLysGluAlaAspPheAspLeu	200	1133	AGTGCACAGGAGTACCGACAGCGAAGCTACTGGAAAAAGAACTGCTTTCTTCGCTTAT	1192
				381	AspValPheGlyIleProPheValAspProAspSerTrpThrProGluGluValIlePro	400
				1193	GATGTTTTTGGAAATTCCTTTGTGATCCGGATTCATGAGCTCCAGAGAAGTATATCCC	1252
				401	LysArgLeuGlnGluLysGlnLysSerGluArgGluThrAlaValArgIleSerGlnGlu	420
				1253	RAGAGTTGCAGAGAGAAACAGAACTCAGACGGGAACAGCCGTACGATCTCCAGGAG	1312
				421	IleGlyAsnLeuMetLysGluIleGluThrLeuValGluGluLysThrLysGluSerLeu	440
				1313	ATTGGAACTTATGAGGAAATCGAGACCTTGTGGAAAGAGAAGACCAAGAGTCACTG	1372
				441	AspValSerArgLeuThrArgGluGlyGlyProLeuLeuTyrGluGlyIleSerLeuThr	460
				1373	GATGAGCAGACATGACCCGGGAAGGTGGCCCCCTGCTGTATGAGGCATCAGTCTCAC	1432
				461	MetAsnSerLysLeuLeuAsnGlyTyrGlnArgValValMetAspGlyValIleSerAsp	480
				1433	ATGAACTCCAAACTCCTGAATGGTTCCCGCGGGTGGTGTGAGCGCGCTAATCTCTGAC	1492
				481	HisGluCysGlnGluLeuGlnArgLeuThrAsnValAlaAlaThrSerGlyAspGlyTyr	500
				1493	CACGAGTGTGAGGAGCTGCAGAGCTGACCAATGTGGCAGCAACCTCAGGAGATGGGTAC	1552
				501	ArgGlyGlnThrSerProHisThrProAsnGluLysPheTyrGlyValThrValPheLys	520
				1553	CGGGGTGAGACTCCCGACATCTCCCAATGAAAGTTCTATGGTGTCTCTCTCAAA	1612
				521	AlaLeuLysLeuGlyGlnGluLysValProLeuGlnSerAlaHisLeuTyrTyrAsn	540
				1613	GCCCTCAAGCTGGGGCAAGAGGCAAAATTCCTCTGAGAGTGGCCACCTGTACTACAAC	1672
				541	ValThrGluLysValArgArgIleMetGluSerTyrPheArgLeuAspThrProLeuTyr	560
				1673	GTGACGGAGAGGTGGCGGCGCATCATGGAGTCTTCTCCGCTTGATGATGATGATGATG	1732

QY 561 PheSerTyrSerHisLeuValCysArgThrAlaIleGluValGlnAlaGluArgLys 580  
Db 1733 TTTTCTACTCTATCTGTTGTCGGCACTGCCATCGAGAGTCCAGGAGGAGAG 1792  
QY 581 AsnAspSerHisProValHisValAlaAspAsnCysIleLeuAsnAlaGluThrLeuValCys 600  
Db 1793 GATGATAGTCACTCAGTCCACGTCGACAACTGCATCTGAATCCCGAGACCTCTCGTGTGT 1852  
QY 601 ValLysGluProProAlaTyrThrPheArgAspTyrSerAlaIleLeuTyrLeuAsnGly 620  
Db 1853 GTCAAGAGCCCCAGCCTACACTTCCGGCATACAGGCCATCTTTTACCTTAATGGG 1912  
QY 621 AspPheAspGlyGlyAsnPheTyrPheThrGluLeuAspAlaLysThrValThrAlaGlu 640  
Db 1913 GACTTCGATGGCGAACTTTTATTTCACTGAATCGGATGCCAAGACCGTGAACGAGAG 1972  
QY 641 ValClnProClnCysGlyArgAlaValGlyPheSerSerGlyThrGluAsnProHisGly 660  
Db 1973 GTGCAGCCTCAGTGTGAAGAGCGGTGGGATCTTTCAGGCACTGAAACCCACATGGA 2032  
QY 661 ValLysAlaValThrArgGlyGlnArgCysAlaIleAlaLeuTyrPheThrLeuAspPro 680  
Db 2033 GTGAGGCTGTCAACAGGGGGCAGCGCTGTGCCATGCCCTGTGTTCACCTGGACCT 2092  
QY 681 ArgHisSerGluArgAspArgValGlnAlaAspAspLeuValLysMetLeuPheSerPro 700  
Db 2093 CGACACAGCGAGCGGACAGGGTGCAGGACATGACCTGTGAGATGCTTTCAGCCCCA 2152  
QY 701 GluGluMetAspLeuSerGlnGluGlnProLeuAspAlaGlnGlnGlyProGluPro 720  
Db 2153 GAGAGATGACCTTCTCCAGAGGACGACCCCTGTGATGCCAGCGGGCCCCCGGACCT 2212  
QY 721 AlaGlnGluSerLeuSerGlySerGlySerGlySerLysProLysAspGluLeu 736  
Db 2213 GCACAGAGTCTCTCTCAGCAGTGAATCGAAGCCCAAGATGAGCTA 2260

RESULT 6  
AAH14513  
ID AAH14513 standard; cDNA; 2524 BP.  
XX AC AAH14513;  
XX DT 26-JUN-2001 (first entry)  
XX DE Human cDNA sequence SEQ ID NO:12044.  
XX KW Human; primer; detection; diagnosis; antisense therapy; gene therapy; ss.  
XX OS Homo sapiens.  
XX PN EP1074617-A2.  
XX PD 07-FEB-2001.  
XX PR 28-JUL-2000; 2000EP-00116126.  
XX PR 29-JUL-1999; 99JP-00248036.  
XX PR 27-AUG-1999; 99JP-00300253.  
XX PR 11-JAN-2000; 2000JP-00118776.  
XX PR 02-MAY-2000; 2000JP-00183767.  
XX PR 09-JUN-2000; 2000JP-00241899.  
XX PA (HELI-) HELIX RES INST.  
XX FI Ota T, Isogai T, Nishikawa T, Hayashi K, Saito K, Yamamoto J;  
PI Ishii S, Sugiyama T, Wakamatsu A, Nagai K, Otsuki T;  
XX WIPI; 2001-318749/34.

PT Primer sets for synthesizing polynucleotides, particularly the 5602 full-length cDNAs defined in the specification, and for the detection and/or diagnosis of the abnormality of the proteins encoded by the full-length

PT cDNAs.  
XX Claim 8; SEQ ID NO 12044; 2537pp + Sequence Listing; English.  
XX The present invention describes primer sets for synthesizing 5602 full-length cDNAs defined in the specification. Where a primer set comprises:  
CC (a) an oligo-dT primer and an oligonucleotide complementary to the  
CC complementary strand of a polynucleotide which comprises one of the 5602  
CC nucleotide sequences defined in the specification, where the  
CC oligonucleotide comprises at least 15 nucleotides; or (b) a combination  
CC of an oligonucleotide comprising a sequence complementary to the  
CC complementary strand of a polynucleotide which comprises a 5'-end  
CC sequence and an oligonucleotide comprising a sequence complementary to a  
CC polynucleotide which comprises a 3'-end sequence, where the  
CC oligonucleotide comprises at least 15 nucleotides and the combination of  
CC the 5'-end sequence/3'-end sequence is selected from those defined in the  
CC specification. The primer sets can be used in antisense therapy and in  
CC gene therapy. The primers are useful for synthesizing polynucleotides,  
CC particularly full-length cDNAs. The primers are also useful for the  
CC detection and/or diagnosis of the abnormality of the proteins encoded by  
CC the full-length cDNAs. The primers allow obtaining of the full-length  
CC cDNAs easily without any specialised methods. AAH03166 to AAH13628 and  
CC AAH13633 to AAH18742 represent human cDNA sequences; AAH92446 to AAH95893  
CC represent human amino acid sequences; and AAH13629 to AAH13632 represent  
CC oligonucleotides, all of which are used in the exemplification of the  
CC present invention  
XX SQ Sequence 2524 BP; 583 A; 747 C; 704 G; 490 T; 0 U; 0 Other;

Alignment Scores:  
Pred. No.: 0 Length: 2524  
Score: 3844.00 Matches: 732  
Percent Similarity: 99.5% Conservative: 1  
Best Local Similarity: 99.46% Mismatches: 3  
Query Match: 99.33% Indels: 0  
DB: 4 Gaps: 0  
US-10-045-815-4 (1-736) x AAH14513 (1-2524)  
QY 1 MetAlaValArgAlaLeuLysLeuLeuThrThrLeuLeuAlaValAlaAlaLys 20  
Db 16 ATGGCGGTACGGCGCTTGAAGCTGCTGACCACTGCTGGGTGTGCTGGCGCTGCCCTCC 75  
QY 21 GlnAlaGluValGluSerGluAlaGlyTyrGlyMetValThrProAspLeuLeuPheAla 40  
Db 76 CAAGCCGAGTTCAGTCCGAGCAGATGGGCGATGTGACGCTGATCTGCTCTCCGCC 135  
QY 41 GluGlyThrAlaAlaTyrAlaArgGlyAspTyrProGlyValValLeuSerMetGluArg 60  
Db 136 GAGGGGACCGCAGCCTACGCGCGGGGACTGGCCCGGGTGGTCTCGAGCATGGAACGG 195  
QY 61 AlaLeuArgSerArgAlaAlaLeuArgAlaLeuArgLeuArgCysArgThrGlnCysAla 80  
Db 196 GCGCTGCGCTCCGGGCGAGCCCTCCGCGCCCTTCCCTTCCGCTGCGCGCACCCAGTGTGCC 255  
QY 81 AlaAspPheProTyrGluLeuAspProAspTyrProSerProAlaGlnAlaSerGly 100  
Db 256 GCCGACTTCCCGTGGAGCTGGACCCGACTGCTCCCGCAGCCGCGCCAGGCTCGGCGC 315  
QY 101 AlaGlyAlaLeuArgAspLeuSerPhePheGlyGlyLeuLeuArgArgAlaAlaCysLeu 120  
Db 316 GCCGCGCGCTCCGCGACCTGAGCTTCTTCGGGGGCTTCTGCGTGGCGCTGCCCTGCC 375  
QY 121 ArgArgCysLeuGlyProProAlaAlaHisSerLeuSerGluGluMetGluLeuGluPhe 140  
Db 376 CGCCGCTGCTCGCGCGCCCGCCGCTGCTGCTCAGGAGAGATGAGCTGAGTTC 435  
QY 141 ArgLysArgSerProTyrAsnTyrLeuGlnValAlaTyrPheLysIleAsnLysLeuGlu 160  
Db 436 CGCAAGCGGAGCGCCCTACAACTACCTGAGGTGCGCTACTTCAAGATCAACAGTTGAG 495  
QY 161 LysAlaValAlaAlaAlaHisThrPheValGlyAsnProGluHisMetGluMetGln 180

496 AAAGCTGTGTCGAGCAGCACACCTTCTCTGTTGGCAATCTGAGCAGACATGGAAATCGAG 555  
181 GlnAsnLeuAspTyrGlnThrMetSerGlyValIysGluAlaAspPheLysAspLeu 200  
556 CAGAACTAGACTATTACCAAAACCATGTCGAGTGAAGGAGGCGGACTTCAAGGATCTT 615  
201 GluThrGlnProHisMetGlnGluPheArgLeuGlyValArgLeuTyrSerGluGluGln 220  
616 GAGACTCAACCCCATATCAAGAATTTGACTGGGAGTGGCACTTACTCAGAGGAACAG 675  
221 ProGlnGluAlaValProHisLeuGluAlaAlaLeuGlnGluTyrPheValAlaTyrGlu 240  
676 CCACAGGAAGCTGTGCCCCCCTAGAGGCGCGCTGCAAGAAATACTTTGTGGCCTATGAG 735  
241 GluCysArgAlaLeuCysGluGlyProTyrAspTyrAspGlyTyrAsnTyrLeuGluTyr 260  
736 GAGTCCCGTGGCCCTCTGCAAGGCGCCCTATGACTACGATGGCTACAACTACTTGAATAC 795  
261 AsnAlaAspLeuPheGlnAlaIleThrAspHisTyrIleGlnValLeuAsnCysLysGln 280  
796 AACCTGACCTTCTCCAGGCCCATACAGATCATTTACATCCAGGTCTCACTGTAAGCAG 855  
281 AsnCysValThrGluLeuAlaSerHisProSerArgGluLysProPheGluAspPheLeu 300  
856 AACTGTGTACGGAGCTTGTCTCCACCCCAAGTCGAGAGAGCCCTTTGAAGACTTCCCTC 915  
301 ProSerHisTyrAsnTyrLeuGlnPheAlaTyrTyrAsnIleGlyAsnTyrThrGlnAla 320  
916 CCATCGCATTAATAATATCTCGAGTTGCTTACTATTAACATTTGGGAATTATACAGGCT 975  
321 GlyGluCysAlaIysThrTyrLeuLeuPhePheProAsnAspGluValMetAsnGlnAsn 340  
976 GTTGAATGTGCAAGACCTATCTTCTCTTCCCAATGACGAGGTGATGAACCAAAAT 1035  
341 LeuAlaTyrTyrAlaAlaMetLeuGlyGluHisThrArgSerIleGlyProArgGlu 360  
1036 TTGGCTTATTATGAGCTATGCTTGGAGAGAAACACACCCAGATCCATTCGCCCCCGGTGAG 1095  
361 SerAlaLysGluTyrArgGlnArgSerLeuLeuGluLysGluLeuLeuPhePheAlaTyr 380  
1096 AGTCCCAAGGATACCGACGAGGCTTACTGGAAGAAAGAACTGCTTTCTTCGCTTAT 1155  
381 AspValPheGlyIleProPheValAspProAspSerTrpThrProGluGluValIlePro 400  
1156 GATCTTTTGGAAATTCCTTTGTGGATCCGATTCATGGACTCCAGAAAGATGATTCCTC 1215  
401 LysArgLeuGlnGluLysGlnLysSerGluArgGluThrAlaValArgIleSerGlnGlu 420  
1216 AAGAGATTGCAAGAGAAACAGAGTCAAGACGGGAAACACCGGTACGCATCTCCAGGAG 1275  
421 IleGlyAsnLeuMethLysGluIleGluThrLeuValGluGluLysThrLysGluSerLeu 440  
1276 ATTGGGAACCTTATGAAGAAATCGAGACCTTGTGGAGAGAAGACCAAGGAGTCACTG 1335  
441 AspValSerArgLeuThrArgGluGlyGlyProLeuLeuTyrGluGlyIleSerLeuThr 460  
1336 GATGTGACGAGCTGACCCGGAAGGTGGCCCCCTGCTGTATGAAGGCACTCAGTCTCACC 1395  
461 MetAsnSerLysLeuLeuAsnGlyTyrGlnArgValValMetAspGlyValIleSerAsp 480  
1396 ATGAACCTCAAACTCTCTGAATGGTTCACGCGGTGTGTGACGCGCTAACTCTGTGAC 1455  
481 HisGlnCysGlnGluLeuGlnArgLeuThrAsnValAlaIleThrSerGlyAspGlyTyr 500  
1456 CACGAGTGTGAGGAGTGCAGAGACTGACCAATGTGGCAGCAACCTCAGAGATGGCTAC 1515  
501 ArgGlyGlnThrSerProHisThrProAsnGlnLysPheTyrGlyValThrValPheLys 520  
1516 CGGGTTCAGACCTCCCACTATCCCAATGAAAGTTCTATGTTGCTCAGTGTCTTCAAA 1575  
521 AlaLeuLysLeuGlyGlnGluGlyLysValProLeuGlnSerAlaHisLeuTyrTyrAsn 540  
1576 GCCTCAAGCTGGGGCAAGAGGCAAGTTCCTCTGACAGTCCCAACCTGTACTACAAC 1635

541 ValThrGluLysValArgIleMetGluSerTyrPheArgLeuAspThrProLeuTyr 560  
1636 GTGAGGAGAGGTGGCGCCATCATGAGTCTCTACTTCCGCTGGATACGCCCTCTAC 1695  
561 PheSerTyrSerHisLeuValCysArgThrAlaIleGluGluValGlnAlaArgLys 580  
1696 TTTTCTCTCTCATCTGTTGGCGGCTGCTCATCGAAGAGGTCCAGCGCAGAGGAAG 1755  
581 AspAspSerHisProValHisValAspAsnCysIleLeuAsnAlaGluThrLeuValCys 600  
1756 GATGATAGTATCATCCAGTCCACGTGCACAACTCATCTCTGAATGCCGACACCTGTGTGT 1815  
601 ValLysGluProProAlaTyrThrPheArgAspTyrSerAlaIleLeuTyrLeuAsnGly 620  
1816 GTCAAGAGCCCCCGGCTTACACCTTCCGCGACTACAGCGCATCTTACTTAATGGG 1875  
621 AspPheAspGlyGlyAsnPheTyrPheThrGluLeuAspAlaLysThrValThrAlaGlu 640  
1876 GACTTCGATGGGGAACCTTTTATTTCTACTGAATGGATGCCAAGACCTGACGCGCAG 1935  
641 ValGlnProGlnCysGlyArgAlaValGlyPheSerSerGlyThrGluAsnProHisGly 660  
1936 GTGCAGCCTCAGTGTGGAAGAGCCGTGGGATTCTCTCAGGCACCTGAAACCCACATGGA 1995  
661 ValLysAlaValThrArgGlyGlnArgCysAlaIleAlaLeuTyrPheThrLeuAspPro 680  
1996 GTGAAGCTGTCCACGAGGGGCGGCTGTGCCATCGCCCTGTGTTCCACCTCGACCCCT 2055  
681 ArgHisSerGluArgAspArgValGlnAlaAspAspLeuValLysMetLeuPheSerPro 700  
2056 CGACACAGCGAGCGGACAGGCTGCGAGCAGATGACCTGTGTGAAGATGCTCTTCAGCCCA 2115  
701 GluGluMetAspLeuSerGlnGluGlnProLeuAspAlaGlnGlnGlyProGluPro 720  
2116 GAAGAGATGGACCTCTCCAGAGCAGCCCATGATGCCAGAGGGCCCCCGAAACCT 2175  
721 AlaGlnGluSerLeuSerGlySerGluSerLysProLysAspGluLeu 736  
2176 GCACAGAGTCTCTCTCAGGCAGTGAATCGAAGCCCAAGGATGAGCTA 2223

RESULT 7  
AAD08489  
ID AAD08489 standard; cDNA; 2577 BP.  
XX  
AC AAD08489;  
XX  
DT 09-AUG-2001 (first entry)  
XX  
DE Human secreted protein-encoding gene 2 cDNA clone HUVFY29, SEQ ID NO:12.  
XX  
KW Human; secreted protein; proliferative disorder; cancer; tumour;  
KW foetal abnormality; developmental abnormality; haematopoietic disorder;  
KW immune system disorder; AIDS; autoimmune disease; rheumatoid arthritis;  
KW inflammation; allergy; neurological disorder; Alzheimer's disease;  
KW Parkinson's disease; cognitive disorder; schizophrenia; asthma;  
KW skin disorder; psoriasis; sepsis; diabetes; kidney disease;  
KW cardiovascular disorder; arginogenic disorder; kidney disease;  
KW gastrointestinal disorder; pregnancy-related disorder; gene therapy;  
KW endocrine disorder; infection; wound healing; vulnerability; cell culture;  
KW chemotaxis; food additive; binding partner identification; ss.  
XX  
OS Homo sapiens.  
XX  
FH Key Location/Qualifiers  
FT 24..1103 /tag= a  
FT /product= "Human secreted protein precursor"  
FT 24..77 /tag= b  
FT /mat\_peptide 78..1100  
FT /product= "Mature human secreted protein"  
FT

XX PN WO200136432-A2.  
XX PD 25-MAY-2001.  
XX PF 15-NOV-2000; 2000WO-US031162.  
XX PR 19-NOV-1999; 99US-0166415P.  
XX PR 30-NOV-2000; 2000US-0215136P.  
XX PA (HUMA-) HUMAN GENOME SCI INC.  
XX PI Ruben SM, Komatsoulis GA, Baker KP, Young PE;  
XX PD P-PSDB; AAE04200.  
XX PT Isolated nucleic acid molecule encoding a human secreted protein is used  
XX PT in preventing, treating or ameliorating a medical condition.  
XX PS Claim 1; Page 403-404; 509pp; English.  
XX CC AAD08488-AA08529 represent cDNAs corresponding to 18 human secreted  
XX CC protein genes, and AAE04199-AAE04239 represent the proteins they encode.  
XX CC AAE04240-AAE04257 represent human secreted protein fragments or variants.  
XX CC The secreted proteins and their genes are useful for preventing, treating  
XX CC or ameliorating medical conditions, e.g., by protein or gene therapy.  
XX CC Pathological conditions can be diagnosed by determining the amount of the  
XX CC new protein in a sample or by determining the presence of mutations in  
XX CC the new genes. Specific uses are described for each of the 18 genes,  
XX CC based on the tissues in which they are most highly expressed, and include  
XX CC developing products for the diagnosis or treatment of proliferative  
XX CC disorders, cancer, tumors, foetal and developmental abnormalities,  
XX CC haematopoietic disorders, diseases of the immune system, AIDS, autoimmune  
XX CC diseases (e.g., rheumatoid arthritis), inflammation, allergies,  
XX CC neurological disorders (e.g., Alzheimer's disease, Parkinson's disease),  
XX CC cognitive disorders, schizophrenia, asthma, skin disorders (e.g.,  
XX CC psoriasis), sepsis, diabetes, atherosclerosis, cardiovascular disorders,  
XX CC angiogenic disorders, kidney disorders, gastrointestinal disorders,  
XX CC pregnancy-related disorders, endocrine disorders, and infections. The  
XX CC proteins can also be used to aid wound healing and epithelial cell  
XX CC proliferation, to prevent skin ageing due to sunburn, to maintain organs  
XX CC before transplantation, for supporting cell culture of primary tissues,  
XX CC to regenerate tissues, to identify their cognate ligands or binding  
XX CC partners, and in chemotaxis, and can be used as a food additive or  
XX CC preservative to modify storage properties. Antibodies specific for a  
XX CC protein of the invention can be used in alleviating symptoms associated  
XX CC with the disorders mentioned above, and in diagnostic immunoassays e.g.,  
XX CC radioimmunoassay or enzyme linked immunosorbent assay (ELISA). The  
XX CC present sequence represents a human secreted protein-encoding cDNA of the  
XX CC invention

XX SQ Sequence 2577 BP; 610 A; 753 C; 711 G; 503 T; 0 U; 0 Other;

## Alignment Scores:

Pred. No.:	0	Length:	2577
Score:	3820.00	Matches:	733
Percent Similarity:	99.32%	Conservative:	0
Best Local Similarity:	99.32%	Mismatches:	3
Query Match:	98.71%	Indels:	2
DB:	4	Gaps:	0

US-10-045-815-4 (1-736) x AAD08489 (1-2577)

Qy	1	MetAlaValArgAlaLeuLysLeuLeuThrThrLeuLeuAlaValAlaAlaSer	20
Db	24	ATGGCGGTACCGCGGTGAAGCTGCTGACCACACTGCTGGCTGTCGGCGCTGCGCTCC	83
Qy	21	GlnAlaGluValGluSerGluAlaGlyTrpGlyMetValThrProAspLeuLeuPheAla	40
Db	84	CAACCCGGGTGCGAGTCCGAGCGAGGATGGGCAATGGTACCGCTGATCTGCTTCGCC	143
Qy	41	GluGlyThrAlaAlaTyrAlaArgGlyAspTrpProGlyValValLeuSerMetGluArg	60

Db	144	GAGGGGACCGCAGCCTACGCGCGGGGACTGGCCCGGGGTGGTCTGAGCATGAACGG	203
Qy	61	AlaLeuArgSerArgAlaAlaLeuArgAlaLeuArgCysArgThrGlnCysAla	80
Db	204	GGCTGGCGCTCCCGGCGAGCCCTCGCGCCCTTTCGCTGCGCTGCCGACCCAGTGTGCC	263
Qy	81	AlaAspPheProTrpGluLeuAspProAspTrpSerProSerProAlaGlnAlaSerGly	100
Db	264	GCCGACTTCCGCTGGGAGCTGGACCCGACTGGTCCCGCAGCCCGCCAGGCTTCGGGC	323
Qy	101	AlaGlyAlaLeuArgAspLeuSerPheGlyGlyLeuLeuArgAlaAlaCysLeu	120
Db	324	GCCGCGCGCTGCGGACCTGAGCTTCTTCGGGGGCGCTTCGCGTGGCGCTGCCTGCGCTG	383
Qy	121	ArgArgCysLeuGlyProProAlaAlaHisSerLeuSerGluGluMetGluLeuPhe	140
Db	384	CGCGCTGCTCGCGCGCGCGCCGCTCGCTCAGCGAAGAGATGGAGTGGAGTTC	443
Qy	141	ArgLysArgSerProTyrAsnTyrLeuGlnValAlaTyrPheLysIleAsnLysLeuGlu	160
Db	444	CGCAAGCGGAGCGCCCTACCACTACCTGCGAGTGGCTTCTTCAAGATCAACAAGTTGGAG	503
Qy	161	LysAlaValAlaAlaAlaHisThrPheValGlyAsnProGluHisMetGluMetGln	180
Db	504	AAAGCTGTGTGTCGACGACACACCTTCTTCGTGGGCAATCTCTGAGCACATGGAATGCGAG	563
Qy	181	GlnAsnLeuAspTyrTyrGlnThrMetSerGlyValLysGluAlaAspPheLysAspLeu	200
Db	564	CAGAACCTAGACTATTACCAACCATGTCTGGAGTGAAGGAGGCGGACTTCAAGGATCTT	623
Qy	201	GluThrGlnProHisMetGlnGluPheArgLeuGlyValArgLeuTyrSerGluGluGln	220
Db	624	GAGACTCAACCCCATGATGCAAGATTCGACTGGGAGTGGGACTCTACTCAGAGGAACAG	683
Qy	221	ProGlnGluAlaValProHisLeuGluAlaLeuGlnGluTyrPheValAlaTyrGlu	240
Db	684	CCACAGGAAGCTGTGCCCCACCTAGAGCGCGCTGCAAGAATACTTTGTGGCTATGAG	743
Qy	241	GluCysArgAlaLeuCysGluGlyProTyrAspTyrAspGlyTyrAsnTyrLeuGluTyr	260
Db	744	GAGTCCGCTGGCCCTCTGGAGGCGCTTATGACTACGATGGCTACAACTTCTGAGTAC	803
Qy	261	AsnAlaAspLeuPheGlnAlaIleThrAspHisTyrIleGlnValLeuAsnCysLysGln	280
Db	804	AACGTGACCTTCTCCAGGCCATACAGATCATATACATCCAGGTCTCACTGTGAAGCAG	863
Qy	281	AsnCysValThrGluLeuAlaSerHisProSerArgGluLysProPheGluAspPheLeu	300
Db	864	AACTGTGTACGAGCTTGTCTCCCAAGTCGAGAGAAGCCCTTTTGAAGACTTCTCTC	923
Qy	301	ProSerHisTyrAsnTyrLeuGlnPheAlaTyrTyrAsnIleGlyAsnTyrThrGlnAla	320
Db	924	CCATCGCATTAATTAATCTGCGAGTTTGCCTACTATAACATTTGGGAATTTATACAGGCT	983
Qy	321	GlyGluCysAlaLysThrTyrLeuLeuPhePheProAsnAspGluValMetAsnGlnAsn	340
Db	984	GTTGAATGTGCCAAGACCTATCTTCTTCTTCCCAATGACGAGTGTATGAACCAAAAT	1043
Qy	341	LeuAla-TyrTyrAlaAlaMetLeuGlyGluLysThrArgSerIleGlyProArgGln	360
Db	1044	TTGGCCCTATTATGAGCTATGCTTGGAGAGAAACACACAGATCCATCGGCCCGCGTA	1103
Qy	360	uSerAlaLysGluTyrArgGlnArgSerLeuLeuGluLysGluLeuLeuPhePheAlaTyr	380
Db	1104	GAGTCCCAAGGAGTACCGACAGCAAGGCTACTTGGAAAAGAACTGCTTTTCTTCGCTTA	1163
Qy	380	rAspValPheGlyIleProPheValAspProAspSerTrpThrProGluGluValIlePr	400
Db	1164	TGATGTTTTTGAATTTCCCTTTGTGGATCCCGATTCATGGACTCCAGAGAGTGAATTC	1223
Qy	400	olysArgLeuGlnGlu-LysGlnLysSerGluArgGluThrAlaValArgIleSerGlnG	420

Db 1224 CAGAGATTGCAAGAGAAACAGAAAGTCAAGACCGGAAACAGCGGTACGATCTCCAGG 1283  
 QY 420 luileGlyAsnLeuMetLysGluIleGluThrLeuValGluGluThrLysGluSerL 440  
 Db 1284 AGATTGGGAACCTTATGAAGAAATCGAGACCCCTGTGTGAAGAGAGACCAAGGAGTCAC 1343  
 QY 440 euAspValSerArgLeuThrArgGluGlyProLeuLeuTyrgluGlyIleSerLeuT 460  
 Db 1344 TGGATGTGAGCAGACTGACCCGGAAGGTGGCCCTGCTGTATGAAGGCAATCAGTCTCA 1403  
 QY 460 hrMetAsnSerLysLeuLeuAsnGlyTyrglnArgValValMetAspGlyValIleSerA 480  
 Db 1404 CCATGAATCTCCAACTCTGAATGGTTCCTCCAGCGGGTGGTGTGATGAGCGGTATCTCTG 1463  
 QY 480 spHisGluCysGlnGluLeuGlnArgLeuThrAsnValAlaAlaThrSerGlyAspGlyT 500  
 Db 1464 ACCACGAGTGTGAGGAGTGCAGAGACTGACCAATGTGGCAGCAACCTCAGGAGATGGCT 1523  
 QY 500 yrArgGlyClnThrSerProHisThrProAsnGluLysPheTyrglyValThrValPheL 520  
 Db 1524 ACCGGGGTCAGACCTCCCAACATCTCCCAATGAAGTTCTATGGTGTCACTGTCTTCA 1583  
 QY 520 ysAlaLeuLysGluGlyGlnGluLysValProLeuGlnSerAlaHisLeuTyrrTyra 540  
 Db 1584 AAGCCCTCAGCTGGGCAAGAGCAAAAGTTCTCTGCGAGTGGCCCACTGTACTACCA 1643  
 QY 540 snValThrGluLysValArgArgileMetGluSerTyrglyPheArgLeuAspThrProLeuT 560  
 Db 1644 ACGTACCGGAGAGTGGCGGCATCATGGAGTCTTCTCCGCTGGATACGCCCTCTCT 1703  
 QY 560 yrPheSerTyrrSerHisLeuValCysArgThrAlaIleGluGluValGlnAlaGluArgL 580  
 Db 1704 ACTTTTCTACTCTCATCTGTGTGGCGCACTGCCATCGAAGAGTCCAGCGAGAGGA 1763  
 QY 580 ysAspAspSerHisProValHisValAspAsnCysIleLeuAsnAlaGluThrIleuValC 600  
 Db 1764 AGGATGATAGTATCCAGTCACGTGACAACTGCATCTCCTGAATGCGGACCCCTGTGT 1823  
 QY 600 ysValLysGluProAlaTyrrPheArgAspTyrrSerAlaIleLeuTyrrLeuAsnG 620  
 Db 1824 GTGTCAAGAGCCCCCAGCTACACCTTCGCGACTACAGCGCATCTTTACCTAAATG 1883  
 QY 620 lyAspPheAspGlyClyAsnPheTyrrPheThrGluLeuAspAlaLysThrValThrAlaG 640  
 Db 1884 GGGACTTCGATGGCGGAACCTTTTATTTCTGTAACCTGGATGCCAAGACCGTGACGCGAG 1943  
 QY 640 luValGlnProGlnCysGlyArgAlaValGlyPheSerSerGlyThrGluAsnProHisG 660  
 Db 1944 AGTGTGACCTCAGTGTGGAAGACCGTGGATCTCTTTCAGGCATGGAACCCACATG 2003  
 QY 660 lyValLysAlaValThrArgGlyGlnArgCysAlaIleAlaLeuTyrrPheThrLeuAspP 680  
 Db 2004 GAGTGAAGGCTGTCCACGAGGGGCGAGCGCTGTGCTGCTGCTGTGCTTCACTGTGACC 2063  
 QY 680 roArgHisSerGluArgAspArgValGlnAlaAspAspLeuValLysMetLeuPheSerP 700  
 Db 2064 CTCGACACAGGCGGACAGGTCAGGTCAGGAGATGACCTGGTGAAGATGCTCTTCAGCC 2123  
 QY 700 roGluGluMetAspLeuSerGlnGluGlnProLeuAspAlaGlnGlnGlyProProGluP 720  
 Db 2124 CAGAAGAGATGACCTCTCCAGGAGCAGCCCTGGATGCCAGCGGGCCCCCCCCGAC 2183  
 QY 720 roAlaGlnGluSerLeuSerGlySerGluSerLysProLysAspGluLeu 736  
 Db 2184 CTGCACAGAGTCTCTCTCAGCGCAGTGAATCGAAGCCCAAGGATGAGCTA 2233

RESULT 8

ABK94910

ID ABK94910 standard; cDNA; 2753 BP.

XX ABK94910;

AC ABK94910;

DT 30-AUG-2002 (first entry)

XX DE Human novel polynucleotide #21.  
 XX KW Human; gene; ss; inflammatory condition; shock; sepsis; immune response;  
 KW cancer; wound healing; central nervous system disease; haematopoiesis;  
 KW peripheral nervous system disease; amyotrophic lateral sclerosis; tendon;  
 KW myeloid cell disorder; lymphoid cell disorder; platelet disorder; bone;  
 KW cartilage; ligament; nerve tissue; ulcer; osteoporosis; osteoarthritis;  
 KW bone degenerative disorder; periodontal disease; reperfusion injury;  
 KW lung fibrosis; liver fibrosis; autoimmune disorder; bacterial infection;  
 KW allergic condition; thrombolysis; thrombosis; coagulation disorder;  
 KW fungal infection.  
 XX OS Homo sapiens.  
 XX FN WO200244340-A2.  
 XX PD 06-JUN-2002.  
 XX 30-NOV-2001; 2001WO-US047004.  
 XX 30-NOV-2000; 2000US-00728952.  
 XX PR (HYSE-) HYSEQ INC.  
 XX PA Tang YT, Goodrich RW, Liu C, Zhou P, Asundi V, Wang J, Wang D;  
 XX PI Yamazaki V, Ujwal ML, Drmanac RT;  
 XX DR WPI; 2002-508509/54.  
 XX DR P-PSDB; ABG66686.  
 XX Novel nucleic acids and polypeptides for diagnosis, treatment of  
 PT inflammatory, autoimmune, nervous system, myeloid or lymphoid cell  
 PT disorders, cancer and promoting wound healing.  
 XX Claim 1; Page 369-373; 672pp; English.

The invention relates to human novel polynucleotides and associated  
 polypeptides. The polynucleotides and polypeptides are useful for  
 treating inflammatory conditions such as arthritis, nephritis, Crohn's  
 disease, ischemia-reperfusion injury, shock, sepsis, immune responses  
 and cancer and for promoting wound healing. The sequences are used to  
 induce the proliferation of neural cells and regeneration of nerve and  
 brain tissue, and are useful for the treatment of central and peripheral  
 nervous system diseases and neuropathies, such as Alzheimer's disease,  
 Parkinson's disease, Huntington's disease and amyotrophic lateral  
 sclerosis. The sequences are involved in chemotactic or chemokinetic  
 activity, regulation of haematopoiesis, treatment of myeloid or lymphoid  
 cell disorders and platelet disorders such as thrombocytopenia,  
 regeneration of bone, cartilage, tendon, ligament and/or nerve tissue  
 growth, tissue repair, healing of burns, incisions, ulcers, treatment of  
 osteoporosis, osteoarthritis, bone degenerative disorders and periodontal  
 disease. The sequences of the invention are also useful for gut  
 protection or regeneration and treatment of lung or liver fibrosis.  
 CC reperfusion injury in various tissues, immune deficiencies and disorders  
 CC including severe combined immunodeficiency (SCID), bacterial or fungal  
 CC infections, autoimmune disorders e.g. multiple sclerosis and myasthenia  
 CC gravis, allergic conditions such as asthma, thrombolysis or thrombosis  
 CC and coagulation disorders. Sequences ABK94890-ABK94982 represent human  
 CC novel polynucleotides of the invention

SQ Sequence 2753 BP; 632 A; 802 C; 784 G; 535 T; 0 U; 0 Other;

Alignment Scores:

Pred. No.:	0	Length:	2753
Score:	3801.00	Matches:	733
Percent Similarity:	90.94%	Conservative:	0
Best Local Similarity:	90.94%	Mismatches:	3
Query Match:	98.22%	Indels:	70
DB:	6	Gaps:	1

US-10-045-815-4 (1-736) x ABK94910 (1-2753)





Db 2161 TTCTCTCAGGCACTGAAACCCACCATGAGTGAAGCTGTACACAGGGGGCAGCGCTGT 2220  
 Qy 671 AlalleAlaLeuTTPreThrLeuAspProArgHisSerGluArgAspArgValGlnAla 690  
 Db 2221 GCCATCGCCCTGTGTTCACCTCGACCTCGACACAGCGAGCGGGACAGGGTGCAGGCA 2280  
 Qy 691 AspAspLeuValLysMetLeuPheSerProGluGluMetAspLeuSerGlnGluPro 710  
 Db 2281 GATGACCTGGTGAAGATGCTCTTCAGCCACAGAGATGAGACTCTCCAGGAGCACCCC 2340  
 Qy 711 LeuAspAlaGlnGlnGlyProProGluProAlaGlnGluSerLeuSerGlySerGluSer 730  
 Db 2341 CTGGATGCCCAAGAGGGGGCCCCCGAACCTGACAAAGAGTCTCTCTCAGGCAAGTATCG 2400  
 Qy 731 LysProLysAspGluLeu 736  
 Db 2401 AAGCCCAAGGATGAGCTA 2418  
 RESULT 9  
 ID ABK94933  
 XX ABK94933 standard; cDNA; 2753 BP.  
 AC ABK94933;  
 XX  
 DT 30-AUG-2002 (first entry)  
 DE Human novel polynucleotide #44.  
 KW Human; gene; ss; inflammatory condition; shock; sepsis; immune response;  
 KW cancer; wound healing; central nervous system disease; haematopoiesis;  
 KW peripheral nervous system disease; amyotrophic lateral sclerosis; tendon;  
 KW myeloid cell disorder; lymphoid cell disorder; platelet disorder; bone;  
 KW cartilage; ligament; nerve tissue; ulcer; osteoporosis; osteoarthritis;  
 KW bone degenerative disorder; periodontal disease; reperfusion injury;  
 KW lung fibrosis; liver fibrosis; autoimmune disorder; bacterial infection;  
 KW allergic condition; thrombolytic; thrombosis; coagulation disorder;  
 KW fungal infection.  
 XX Homo sapiens.  
 OS  
 XX WO200244340-A2.  
 PN  
 XX 06-JUN-2002.  
 PD  
 XX 30-NOV-2001; 2001WO-US047004.  
 PF  
 XX 30-NOV-2000; 2000US-00728952.  
 PR  
 XX (HYSE-) HYSEQ INC.  
 PA  
 XX Tang YT, Goodrich RW, Liu C, Zhou P, Asundi V, Wang J, Wang D;  
 PI Yamazaki V, Ujwal ML, Drmanac RT;  
 XX  
 XX WPI; 2002-508509/54.  
 DR P-PSDB; ABG66709.  
 XX  
 XX Novel nucleic acids and polypeptides for diagnosis, treatment of  
 PT inflammatory, autoimmune, nervous system, myeloid or lymphoid cell  
 PT disorders, cancer and promoting wound healing.  
 XX  
 XX Claim 1; Page 441-444; 672pp; English.

CC cell disorders and platelet disorders such as thrombocytopenia,  
 CC regeneration of bone, cartilage, tendon, ligament and/or nerve tissue  
 CC growth, tissue repair, healing of burns, incisions, ulcers, treatment of  
 CC osteoporosis, osteoarthritis, bone degenerative disorders and periodontal  
 CC disease. The sequences of the invention are also useful for gut  
 CC protection or regeneration and the invention of lung or liver fibrosis,  
 CC reperfusion injury in various tissues, immune deficiencies and disorders  
 CC including severe combined immunodeficiency (SCID), bacterial or fungal  
 CC infections, autoimmune disorders e.g. multiple sclerosis and myasthenia  
 CC gravis, allergic conditions such as asthma, thrombolytic or thrombotic  
 CC and coagulation disorders. Sequences ABK94890-ABK94982 represent human  
 CC novel polynucleotides of the invention  
 XX  
 SQ Sequence 2753 BP; 632 A; 802 C; 784 G; 535 T; 0 U; 0 Other;  
 Alignment Scores:  
 Pred. No.: 0 Length: 2753  
 Score: 3801.00 Matches: 733  
 Percent Similarity: 90.94% Conservative: 0  
 Best Local Similarity: 90.94% Mismatches: 3  
 Query Match: 98.22% Indels: 70  
 DB: 6 Gaps: 1  
 US-10-045-815-4 (1-736) x ABK94933 (1-2753)  
 QY 1 MetAlaValArgAlaLeuLysLeuLeuThrThrLeuLeuAlaValAlaAlaSer 20  
 DB 1 ATGGCGGTACGGCGCTTGAAGCTGCTGACACACACTGCTGGCTGTGGCGCTGCTCC 60  
 QY 21 GlnAlaGluValGluSerGluAlaGlyTrpGlyMetValThrProAspLeuPheAla 40  
 DB 61 CAAGCCGAGGTGAGTCCGAGCGAGATGGGCGATGAGCGCTGATCTCTTTCGCC 120  
 QY 41 GluGlyThrAlaAlaTyAlaArgGlyAspTrpProGlyValValLeuSerMetGluArg 60  
 DB 121 GAGGGAGCCGAGCCTACGCGCGGGGACTGGCCCGGGGTGGTCTTGAGCATGGACGG 180  
 QY 61 AlaLeuArgSerArgAlaAlaLeuArgAlaLeuArgLeuArgCysArgThrGlnCysAla 80  
 DB 181 GCGCTGCGCTCCCGGGCAGCCCTCGCGCCCTTGGCTGCGCTCCGCGCACCGATGTGCC 240  
 QY 81 AlaAspPheProTrpGluLeuAspProAspTrpSerProSerProAlaGlnAlaSerGly 100  
 DB 241 GCGGATTCCTCGTGGAGCTGGACCCCGACTGGTCCCCCAGCCCGCCCGCTCGGGC 300  
 QY 101 AlaGlyAlaLeuArgAspLeuSerPheGlyGlyLeuLeuArgAlaAlaCysLeu 120  
 DB 301 GCGCGCGCTCGCGACCTGAGCTTCTTCGGGGGCGCTTCTGCGTGGCGCTGCGCTG 360  
 QY 121 ArgArgCysLeuGlyProProAlaAlaHisSerLeuSerGluMetGluLeuGluPhe 140  
 DB 361 CCGCCCTGCTCGGGCCCGCCCGCCCTCGCTCAGCGAGAGATGGAGCTGGAGTTC 420  
 QY 141 ArgLysArgSerProTyAsnTyxLeuGlnValAlaTyPhe----- 154  
 DB 421 CGCAAGCGAGCGCCCTACAACTACTCGTGGAGTGGCTACTTCAAGGTGCAGACCTGCTG 480  
 QY 154 ----- 154  
 DB 481 GAACAGCGCGCGGGGTCTTCTTGGGGAGAGAGTGTTCGAGGGGAGCTGAGAGCTTG 540  
 QY 154 ----- 154  
 DB 541 GGGGATCGGGAGAGTGTCCGCGAGGGAGGGGAAAGTGGCTCTCTGCTGGGAGCTCTCT 600  
 QY 154 ----- 154  
 DB 601 CGGAGCCGGGAGAGTGTCTCTCGGAGGAGAGCTTCTCTCGCCAGTTCGATGGGAG 660  
 QY 155 -----LysileAsnLysLeuGluLysAlaValAlaAlaHisThrPhePhe 170  
 DB 661 ATGCTAACCCCAAGATCAACAAGTTGGAGAAAGCTGTGTCTGCAGCACACACCTTCTTC 720



171 ValGlyAsnProGluHisMetGluMetGlnAsnLeuAspTyrTyrGlnThrMetSer 190  
172 GTGGCAATCTGAGCACATGGAATGCAGAGAACCTGAGCTATTACCAACCATGTCT 780  
191 GlyValLysGluAlaAspPheLysAspLeuGluThrGlnProHisMetGlnGluPheArg 210  
192 GTGGTGAAGAGGCGGACTTCAGAGATCTTGAGACTCAACCCCATATGCAAGAATTTCGA 840  
211 LeuGlyValArgLeuTyrSerGluGluGlnProGlnGluAlaValProHisLeuGluAla 230  
212 CTGGGAGTGGGACTCTACTCAGAGGAACAGCCACAGGAAGCTGTGCCCCACCTAGAGCG 900  
231 AlaLeuGlnGluTyrPheValAlaTyrGluGluCysArgAlaLeuCysGluGlyProTyr 250  
232 GCGCTGCAAGAATACTTTGTGGCCATGAGAGTGGCGTCCCTCTGCGAAGGGCCCTAT 960  
251 AspTyrAspGlyTyrAsnTyrLeuGluTyrAsnAlaAspLeuPheGlnAlaLeuThrAsp 270  
252 GACTACGATGGCTACACTACCTTGTAGTACAACTGACCTCTCCAGGCCATCACAGAT 1020  
271 HisTyrLeuGlnValLeuAsnCysLysGlnAsnCysValThrGluLeuAlaSerHisPro 290  
272 CATTACATCCAGGTCCTCAACTGTAAGCAGAACTGTCTCAGGAGCTTGCTTCCACCCA 1080  
291 SerArgGluLysProPheGluAspPheLeuProSerHisTyrAsnTyrLeuGlnPheAla 310  
292 AGTCGAGAGAGCCCTTTGAGACTTCTCCCATCGCATTAATTAATCTGCAGTTTGGC 1140  
311 TyrTyrAsnLeuGlyAsnTyrThrGlnAlaGlyCysAlaLysThrTyrLeuLeuPhe 330  
312 TACTATAACATTGGGAATTATACACAGCTGTGTAATGTGCCAAGACTATCTTCTTTC 1200  
331 PheProAsnAspGluValMetAsnGlnAsnLeuAlaTyrTyrAlaAlaMetLeuGlyGlu 350  
332 TTCCCAATCAGAGGTGATGAACAAAATTTGGCCCTATTATGAGCTATGCTTTGGAGAA 1260  
351 GluHisThrArgSerIleGlyProArgGluSerAlaLysGluTyrArgGlnArgSerLeu 370  
352 GAACACACCATCATCGGCCCTCGTAGAGTGCCAGAGGTACCGACAGCGAGCCTTA 1320  
371 LeuGluLysGluLeuLeuPhePheAlaTyrAspValPheGlyIleProPheValAspPro 390  
372 CTGGAAAAGAACTGCTTTTCTCGCTTATGATGTTTGGAAATTTCCCTTTGTGGATCCG 1380  
391 AspSerTrpThrProGluGluValIleProLysArgLeuGlnGluLysGlnLysSerGlu 410  
392 GATTCACTGGATCCAGAGAAAGTATTTCCCAAGAGATTGCAAGAAAACAGAGTCAAGAA 1440  
411 ArgGluThrAlaValArgIleSerGlnGluIleGlyAsnLeuMetLysGluIleGluThr 430  
412 CGGGAACACAGCCGTACGCATCTCCAGAGATTGCGAACTTATGAAGGAATTCGAGACC 1500  
431 LeuValGluGluLysThrLysGluSerLeuAspValSerArgLeuThrArgGluGlyGly 450  
432 CTTGTGAAGAGAAAGACCAAGAGTCTACTGATGTGAGCAGACTGACCGGGAAGGTGGC 1560  
451 ProLeuLeuTyrGluGlyIleSerLeuThrMetAsnSerLysLeuLeuAsnGlyTyrGln 470  
452 CCCCTGTGTATGAGGCATCAGTCTCACCATGAACTCCAACTCCTGATGTTCCAG 1620  
471 ArgValValMetAspGlyValIleSerAspHisGluCysGlnGluLeuGlnArgLeuThr 490  
472 CGGGTGTGTATGAGCGCGTAACTCTGACCACAGTGTCTCAGGAGCTGTCAGAGACTGACC 1680  
491 AsnValAlaAlaThrSerGlyAspGlyTyrArgGlyGlnThrSerProHisThrProAsn 510  
492 AATGTGGAGCAACTCAGAGATGGCTACCGGGGTGAGACTCTCCCACTACTCCCAAT 1740  
511 GluLysPheTyrGlyValThrValPheLysAlaLeuLysLeuGlyGlnGluLysVal 530  
1741 GAAAAGTTCTATGTGTCTACTGTCTTCAAGCCCTCAAGCTGGGGCAAGAGCAAGTT 1800  
531 ProLeuGlnSerAlaHisLeuTyrTyrAsnValThrGluLysValArgGlnLeuMetGlu 550

1801 CCTCTGCAGAGTGCACCTGTACTACAACGTGACGGAAGGTGCGCGCATCATGAG 1860  
551 SerTyrPheArgLeuAspThrProLeuTyrPheSerTyrSerHisLeuValCysArgThr 570  
1861 TCCTACTTCCGCTCGATACGCCCTCTACTTTCTACTCTCATCTGGTGGCGCACT 1920  
571 AlaIleGluGluValGlnAlaGluArgLysAspSerHisProValHisValAspAsn 590  
1921 GCCATCGAAGAGTCCAGGACAGAGAGGATGATGATCATCTCCAGTCCACGTGACAC 1980  
591 CysIleLeuAsnAlaGluThrLeuValCysValLysGluProAlaTyrThrPheArg 610  
1981 TGCACTCTGAATCCGAGACCTCTGTGTGTCAAGAGCCGCCAGCTTACACTTCGCG 2040  
611 AspTyrSerAlaIleLeuTyrLeuAsnGlyAspPheAspGlyGlyAsnPheTyrPheThr 630  
2041 GACTACAGCCCATCTTTTACCTAAATGGGACTTCGATGGCGGAACTTTTATTTCAT 2100  
631 GluLeuAspAlaLysThrValThrAlaGluValGlnProGlnCysGlyArgAlaValGly 650  
2101 GAACTGGATGCAAGACCGTGACGAGAGGTGACGCTCAGTGTGGAAGAGCCGTGGGA 2160  
651 PheSerSerGlyThrGluAsnProHisGlyValLysAlaValThrArgGlyGlnArgCys 670  
2161 TTCTCTTCAGGCACTGAAAACCCACATGGAGTGAAGGTGTCCAGGGGGCAGCGCTGT 2220  
671 AlaIleAlaLeuTyrPheThrLeuAspProArgHisSerGluArgAspArgValGlnAla 690  
2221 GCCATCGCCCTGTGTTCACCTGACCTGCACACGACGAGCGGACAGGGGTGAGGCA 2280  
691 AspAspLeuValLysMetLeuPheSerProGluGluMetAspLeuSerGlnGluPro 710  
2281 GATGACCTGTGGAAGATGCTCTTCAGCCCAAGAGATGAGACTCTCCAGAGGAGAGCC 2340  
711 LeuAspAlaGlnGlnGlyProProGluProAlaGlnGluSerLeuSerGlySerGluSer 730  
2341 CTGGATGCCACAGCGGCCCGCCCGAACCTGCACAAAGAGTCTCTCTCAGGCACTGAATCG 2400  
731 LysProLysAspGluLeu 736  
2401 AAGCCCAAGGATGAGCTA 2418  
RESULT 10  
AAD08507  
ID AAD08507 standard; cDNA; 2615 BP.  
XX  
AC AAD08507;  
DT 09-AUG-2001 (first entry)  
XX  
DE Human secreted protein-encoding gene 2 cDNA clone HUVFY29, SEQ ID NO:30.  
XX  
KW Human; secreted protein; proliferative disorder; cancer; tumour;  
KW foetal abnormality; developmental abnormality; haematopoietic disorder;  
KW immune system disorder; AIDS; autoimmune disease; rheumatoid arthritis;  
KW inflammation; allergy; neurological disorder; Alzheimer's disease;  
KW Parkinson's disease; cognitive disorder; schizophrenia; asthma;  
KW skin disorder; psoriasis; sepsis; diabetes; atherosclerosis;  
KW cardiovascular disorder; angiogenic disorder; kidney disorder;  
KW gastrointestinal disorder; pregnancy-related disorder; gene therapy;  
KW endocrine disorder; infection; wound healing; vunerary; cell culture;  
KW chemotaxis; food additive; binding partner identification; ss.  
XX  
OS Homo sapiens.  
XX  
XX  
PH Key Location/Qualifiers  
FT CDS 24..545  
FT /tag= a  
FT /product= "Human secreted protein precursor"  
FT sig\_peptide 24..77  
FT /tag= b  
FT mat\_peptide 78..542

FT /\*tag= C  
 XX /product= "Mature human secreted protein"  
 PN WO200136432-A2.

XX 25-MAY-2001.

XX 15-NOV-2000; 2000WO-US031162.

XX 19-NOV-1999; 99US-0166415P.

XX 30-JUN-2000; 2000US-0215136P.

XX (HUMA-) HUMAN GENOME SCI INC.

XX Ruben SM, Komatsoulis GA, Baker KP, Young PE;

XX P-PSDB; AAE04217.

XX WPI; 2001-343793/36.

XX Isolated nucleic acid molecule encoding a human secreted protein is used

PT in preventing, treating or ameliorating a medical condition.

XX Claim 1; Page 416-417; 509pp; English.

CC AAD08488-AAD08529 represent cDNAs corresponding to 18 human secreted  
 CC protein genes, and AAE04199-AAE04239 represent the proteins they encode.  
 CC AAE04240-AAE04297 represent human secreted protein fragments or variants.  
 CC The secreted proteins and their genes are useful for preventing, treating  
 CC or ameliorating medical conditions, e.g., by protein or gene therapy.  
 CC Pathological conditions can be diagnosed by determining the amount of the  
 CC new protein in a sample or by determining the presence of mutations in  
 CC the new genes. Specific uses are described for each of the 18 genes,  
 CC based on the tissues in which they are most highly expressed, and include  
 CC developing products for the diagnosis or treatment of proliferative  
 CC disorders, cancer, tumors, foetal and developmental abnormalities,  
 CC haematopoietic disorders, diseases of the immune system, AIDS, autoimmune  
 CC diseases (e.g., rheumatoid arthritis), inflammation, allergies,  
 CC neurological disorders (e.g., Alzheimer's disease, Parkinson's disease),  
 CC cognitive disorders, schizophrenia, asthma, skin disorders (e.g.,  
 CC psoriasis), sepsis, diabetes, atherosclerosis, cardiovascular disorders,  
 CC anogenic disorders, kidney disorders, gastrointestinal disorders,  
 CC pregnancy-related disorders, endocrine disorders, and infections. The  
 CC proteins can also be used to aid wound healing and epithelial cell  
 CC proliferation, to prevent skin ageing due to sunburn, to maintain organs  
 CC before transplantation, for supporting cell culture of primary tissues,  
 CC to regenerate tissues, to identify their cognate ligands or binding  
 CC partners, and in chemotaxis, and can be used as a food additive or  
 CC preservative to modify storage properties. Antibodies specific for a  
 CC protein of the invention can be used in alleviating symptoms associated  
 CC with the disorders mentioned above, and in diagnostic immunoassays e.g.,  
 CC radioimmunoassay or enzyme linked immunosorbent assay (ELISA). The  
 CC present sequence represents a human secreted protein-encoding cDNA of the  
 CC invention

SQ Sequence 2615 BP; 616 A; 764 C; 717 G; 514 T; 0 U; 4 Other;

#### Alignment Scores:

Pred. No.:	0	Length:	2615
Score:	3792.00	Matches:	731
Percent Similarity:	99.32%	Conservative:	0
Best Local Similarity:	99.32%	Mismatches:	5
Query Match:	97.98%	Indels:	2
DB:	4	Gaps:	0

US-10-045-815-4 (1-736) x AAD08507 (1-2615)

Qy 1 MetAlaValArgAlaLeuLysLeuThrThrLeuLeuAlaValAlaAlaSer 20

Db 24 ATGGCGGTACGGCGTTGAAGTGTGACACACTGTGCTGTGCTGTGCGCGGTGCTCC 83

Qy 21 GlnAlaGluValGluSerGluAlaGlyTyrGlyMetValThrProAspLeuPheAla 40

Db 84 CAAGCCGAGGTTCGAGTCCGAGGAGGATGGGATGGTGTGACGCTGATCTGCTCTTCGCG 143

Qy	41	GluGlyThrAlaAlaValArgGlyAspTrpProGlyValValLeuSerMetGluArg	60
Db	144	GAGGGGACCGCAGCCTACGCGCGGGGACTGGCCCGGGTGGTCTCTGAGCATGGACGG	203
Qy	61	AlaLeuArgSerArgAlaAlaLeuArgAlaLeuArgCysArgThrGlnCysAla	80
Db	204	GGGCTGGCTCCCGGGGACGCCCTCCGCGCCCTTCGCTGGCTGGCTGGCCACCGAGTGTCC	263
Qy	81	AlaAspPheProTrpGluLeuAspProAspTrpSerProSerProAlaGlnAlaSerGly	100
Db	264	GCCGACTTCCCGTGGAGCTGGACCCCGACTGGTCCCCCAGCGCCGCGCCCTCGGGC	323
Qy	101	AlaGlyAlaLeuArgAspLeuSerPheGlyGlyLeuLeuArgArgAlaAlaCysLeu	120
Db	324	GCGCGCGCTCGCGGACCTGAGCTTCTTGGGGGCGCTTCTGCGTGGCTGCGCTGCGCTG	383
Qy	121	ArgArgCysLeuGlyProProAlaAlaHisSerLeuSerGluGluMetGluLeuGluPhe	140
Db	384	CGCCGCTGCTCGGCG	442
Qy	141	ArgLysArgSerProTyrAsnTyrLeuGlnValAlaTyrPheLysIleAsnLysLeuGlu	160
Db	443	CGCAAGCGGAC-CCCTACAACTTACCTGCAGGTGCGCTACTTCAAGATCAACAAGTTGGAG	501
Qy	161	LysAlaValAlaAlaAlaHisThrPheValGlyAsnProGluHisMetGluMetGln	180
Db	502	AAAGCTGTGTGCGACACACACCTTCTTGTGGCAATCTCTGAGCATATGGAATGGAG	561
Qy	181	GlnAsnLeuAspTyrTyrGlnThrMetSerGlyValLysGluAlaAspPheLysAspLeu	200
Db	562	CAGAACCTAGACTATTACCAACCATGCTGGAGTGAAGGAGGCGGCGCTTCAAGGATCTT	621
Qy	201	GluThrGlnProHisMetGlnGluPheArgGluGlyValArgLeuTyrSerGluGluGln	220
Db	622	GAGACTCAACCCCATATCAAGAAATTTGACTGGGAGTGGGAGTCTACTCAGAGGAACAG	681
Qy	221	ProGlnGluAlaValProHisLeuGluAlaAlaLeuGlnGluTyrPheValAlaTyrGlu	240
Db	682	CCACAGGAGGCTGTGCCCCACCTAGAGCGGGCTGCAAGATATTTGTGGCTATGAG	741
Qy	241	GluCysArgAlaLeuCysGluGlyProTyrAspTyrAspGlyTyrAsnTyrLeuGluTyr	260
Db	742	GAGTCCCGTGGCTCTGCGAAGGGCCCTATGACTACGATGGCTACCACTACTTGGAGTAC	801
Qy	261	AsnAlaAspLeuPheGlnAlaIleThrAspHisTyrIleGlnValLeuAsnCysLysGln	280
Db	802	AACGCTGACCTTCTCCAGGCCATCACAGATCATTACATCCAGGTCCTCACTGTAGAGAG	861
Qy	281	AsnCysValThrGluLeuAlaSerHisProSerArgGluLysProPheGluAspPheLeu	300
Db	862	AACTGTGTCAAGGAGCTTGTCTCCACCAAGTCGAGAGAAGCCCTTTGAAGACTTCCTC	921
Qy	301	ProSerHisTyrAsnTyrLeuGlnPheAlaTyrTyrAsnIleGlyAsnTyrGlnAla	320
Db	922	CCATCGCATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA	981
Qy	321	GlyGluCysAlaLysThrTyrLeuLeuPhePheProAsnAspGluValMetAsnGlnAsn	340
Db	982	GTTGAATGTCCAGACACCTACT	1041
Qy	341	LeuAlaTyrTyrAlaAlaMetLeuGlyGluGluHisThrArgSerIleGlyProArgGlu	360
Db	1042	TTGGCTCTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT	1101
Qy	361	SerAlaLysGluTyrArgGlnArgSerLeuLeuLysGluLeuLeuPhePheAlaTyr	380
Db	1102	AGTGCCAGGAGTACCGACACGGAAGCCTACTGGAAGAAAGAACTGCTTTCTTCGCTTAT	1161
Qy	381	AspValPheGlyIleProPheValAspProAspSerTrpThrProGluGluValIlePro	400
Db	1162	GATGTTTTTGAATTCCTTTGTGATCCCGGATTCATGGACTCCAGAGAAGTGAATTCCTC	1221

Qy 401 LysArgLeuGlnGlnLysGlnLysSerGluArgGluThrAlaValArgIleSerGlnGlu 420  
Db 1222 AAGAGATTGCAAGAGAAACAGAGTTCAGAGCGGAAACAGCGTATCCAGGAG 1281  
Qy 421 IleGlyAsnLeuMetLysGlnIleGluThrLeuValGluGlyThrLysGluSerLeu 440  
Db 1282 ATTGGGAACCTTATGAGGAAATCGAGACCTTGTGGAAGAGAACCAAGGAGTCACTG 1341  
Qy 441 AspValSerArgLeuThrArgGluGlyGlyProLeuLeuTyrGluGlyIleSerLeuThr 460  
Db 1342 GATGTGAGCAGATGATCCCGGAGAGTGGCCCTGCTGATGAAGCATCAGTCTCACC 1401  
Qy 461 MetAsnSerLysLeuLeuAsnGlyTyrGlnArgValValMetAspGlyValIleSerAsp 480  
Db 1402 ATGAACCTCCAAACTCTCTGAATGTTCCAGCGGTGTGATGACGCGCTAATCTCTGAC 1461  
Qy 481 HisGluCysGlnGlnLeuGlnArgLeuThrAsnValAlaAlaThrSerGlyAspGlyTyr 500  
Db 1462 CAGAGTGTGAGAGCTGACAGACTGACCAATGTGGACCACTTCAGAGATGGCTAC 1521  
Qy 501 ArgGlyGlnThrSerProHisThrProAsnGlnLysPheTyrGlyValThrValPheLys 520  
Db 1522 CGGGGTGAGACCTCCGCACATCTCCCAATGAAAGTTCATGGTGTCTACTGTCTTCAA 1581  
Qy 521 AlaLeuLysLeuGlnGlnGlyLysValProLeuGlnSerAlaHisLeuTyrTyrAsn 540  
Db 1582 GCCTCTAAGCTGGGCAAGAGCAAGATTCTCTGAGAGTGGCCACCTGTACTACAAC 1641  
Qy 541 ValThrGluLysValArgArgIleMetGluSerTyrPheArgLeuAspThrProLeuTyr 560  
Db 1642 GTGACGGAGAGGTGGCGGCATCATGAGTCTCTACTTCGCTCGATAGCCCTCTAC 1701  
Qy 561 PheSerTyrSerHisLeuValCysArgThrAlaIleGluValGlnAlaGluArgLys 580  
Db 1702 TTTTCTACTCTCATCTGGTGTGCGCACTGCCATCGAAGAGTCCAGGCAGAGAGAG 1761  
Qy 581 AspAspSerHisProValHisValAspAsnCysIleLeuAsnAlaGluThrLeuValCys 600  
Db 1762 GATGATAGTATCATGATCCAGTGGACACTGATCTCTGATGCCAGACCTCTGTGTGT 1821  
Qy 601 ValLysGluProProLafThrPheArgAspTyrSerAlaIleLeuTyrLeuAsnGly 620  
Db 1822 GTCAAG 1881  
Qy 621 AspPheAspGlyGlyAsnPheTyrPheThrGluLeuAspAlaLysThrValThrAlaGlu 640  
Db 1882 GACTTCGATCGCGAAACTTTTATTTTCACTGAACTGGATGCCAAGACCGTGCAGGAG 1941  
Qy 641 ValGlnProGlnCysGlyArgAlaValGlyPheSerSerGlyThrGluAsnProHisGly 660  
Db 1942 GTCAAG 2001  
Qy 661 ValLysAlaValThrArgGlyGlnArgCysAlaIleAlaLeuThrPheThrLeuAspPro 680  
Db 2002 GTGAAGGCTGTCAAG 2061  
Qy 681 ArgHisSerGluArgAspArgValGlnAlaAspLeuValLysMetLeuPheSerPro 700  
Db 2062 CGACACAG 2121  
Qy 701 GluGluMetAspLeuSerGlnGlnProLeuAspAlaGlnGlnGlyProGluPro 720  
Db 2122 GAAGAGATGACCTCTCTCCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2181  
Qy 721 AlaGlnGluSerLeuSerGlySerGluSerLysProLysAspGluLeu 736  
Db 2182 GCACAGAGAGTCTCTCTCAGGAGAGTGAATCGAAGCCCAAGAGATGAGCTA 2229

RESULT 11  
AAC64724  
ID AAC64724 standard; cDNA; 2829 BP.  
XX  
AC AAC64724;

XX 27-FEB-2001 (first entry)  
DT Human tumor suppressor Gros1-L encoding cDNA SEQ ID NO:1.  
DE  
XX Human tumor suppressor Gros1-L, Gros1-S; cell proliferation; regulation;  
XX Tumor suppressor; Gros1-L; Gros1-S; cell proliferation; regulation;  
KW cancer; cytostatic; Gene therapy; ss.  
XX  
OS Homo sapiens.  
XX WO2000065047-A1.  
XX  
PD 02-NOV-2000.  
XX  
PF 26-APR-2000; 2000WO-JP002731.  
XX  
PR 26-APR-1999; 95JP-00118806.  
XX (CHUG-) CHUGAI RES INST MOLECULAR MEDICINE INC.  
PA  
XX Wadhwa R, Sugihara T, Yoshida A;  
XX WPI; 2000-687340/67.  
XX P-PSDB; AAB36391.  
XX Full-length tumor suppressor genes encoding Gros1-L, Gros1-S and mouse  
PT homologs participating in regulation of cell proliferation, useful in  
PT development of preventives and remedies of cancer.  
XX  
PS Claim 1; Page 61-66; 114pp; Japanese.  
XX  
CC The present sequence encodes the human tumor suppressor designated Gros1  
CC -L. Gros1-L and Gros1-S have cytostatic activity and can be used in gene  
CC therapy. Gros1-L and Gros1-S genes are useful in the development of drugs  
CC used to treat and prevent cancer  
XX  
SQ Sequence 2829 BP; 659 A; 808 C; 801 G; 561 T; 0 U; 0 Other;

Alignment Scores:  
Pred. No.: 0 Length: 2829  
Score: 3780.00 Matches: 733  
Percent Similarity: 90.41% Conservative: 2  
Best Local Similarity: 90.16% Mismatches: 1  
Query Match: 97.67% Indels: 77  
DB: 3 Gaps: 2  
US-10-045-815-4 (1-736) x AAC64724 (1-2829)

Qy 1 MetAlaValArgAlaLeuLysLeuThrThrLeuLeuAlaValAlaAlaAlaSer 20  
Db 52 ATGGCGGTACGCGGTGAAGCTGCTGACACACTGCTGGCTGCTGCGCGCTCC 111  
Qy 21 GlnAlaGluValGluSerGluAlaGlyTrpGlyMetValThrProAspLeuPheAla 40  
Db 112 CAAGCCGAGTGCAGTCCGAGCAGGATGGGCGCATGGTGACCGCTGATCTGCTTCGCC 171  
Qy 41 GluGlyThrAlaAlaTyrAlaArgGlyAspTrpProGlyValValLeuSerMetGluArg 60  
Db 172 GAGGGGACCGGACCTACGCGCGGGGACTGGCCCGGGGTGGTCTCTGAGCATGAAACGG 231  
Qy 61 AlaLeuArgSerArgAlaAlaLeuArgAlaLeuArgLeuArgCysArgThrGlnCysAla 80  
Db 232 GCGCTGCGCTCCCGGCGAGCCCTCGCGCCCTTGGCTGGCTGGCGCAGCCAGGTGCC 291  
Qy 81 AlaAspPheProTrpGluLeuAspProAspTrpSerProSerProAlaGlnAlaSerGly 100  
Db 292 GCGGACTTCCCGTGGAGCTGACCCCGACTGGTCCCCCAGCCCGCCAGCCCTCGGGC 351  
Qy 101 AlaGlyAlaLeuArgAspLeuSerPheGlyGlyLeuLeuArgAlaAlaCysLeu 120  
Db 352 GCGCGCGCCCTCGCGAGCTTCTTTCGGGGGCTTCTGCGTGGCTGCTGCTGCTGCTG 411  
Qy 121 ArgArgCysLeuGlyProProAlaAlaHisSerLeuSerGluGluMetGluLeuPhe 140



XX Human cDNA sequence SEQ ID NO:15159.  
XX Human; primer: detection; diagnosis; antisense therapy; gene therapy; ss.  
XX Homo sapiens.  
XX EF1074617-A2.  
XX 07-FEB-2001.  
XX 28-JUL-2000; 2000EP-00116126.  
XX 29-JUL-1999; 99JP-00248036.  
XX 27-AUG-1999; 99JP-00300253.  
XX 11-JAN-2000; 2000JP-00118776.  
XX 02-MAY-2000; 2000JP-00183767.  
XX 09-JUN-2000; 2000JP-00241899.  
XX (HELI-) HELIX RES INST.  
XX Ota T, Isogai T, Nishikawa T, Hayashi K, Saito K, Yamamoto J;  
XX Ishii S, Sugiyama T, Wakamatsu A, Nagai K, Otsuki T;  
XX WPI; 2001-318749/34.  
XX Primer sets for synthesizing polynucleotides, particularly the 5602 full-length cDNAs defined in the specification, and for the detection and/or diagnosis of the abnormality of the proteins encoded by the full-length cDNAs.  
XX Claim 8; SEQ ID NO 15159; 2537pp + Sequence Listing; English.  
XX The present invention describes primer sets for synthesizing 5602 full-length cDNAs defined in the specification. Where a primer set comprises: (a) an oligo-dT primer and an oligonucleotide complementary to the complementary strand of a polynucleotide which comprises one of the 5602 nucleotide sequences defined in the specification, where the oligonucleotide comprises at least 15 nucleotides; or (b) a combination of an oligonucleotide comprising a sequence complementary to the complementary strand of a polynucleotide which comprises a 5'-end sequence and an oligonucleotide comprising a sequence complementary to a polynucleotide which comprises a 3'-end sequence, where the oligonucleotide comprises at least 15 nucleotides and the combination of the 5'-end sequence/3'-end sequence is selected from those defined in the specification. The primer sets can be used in antisense therapy and in gene therapy. The primers are useful for synthesizing polynucleotides, particularly full-length cDNAs. The primers are also useful for the detection and/or diagnosis of the abnormality of the proteins encoded by the full-length cDNAs. The primers allow obtaining of the full-length cDNAs easily without any specialised methods. AAH03166 to AAH13628 and AAH13633 to AAH18742 represent human cDNA sequences; AAB92446 to AAB95893 represent human amino acid sequences; and AAH13629 to AAH13632 represent oligonucleotides, all of which are used in the exemplification of the present invention  
XX SQ Sequence 2993 BP; 666 A; 864 C; 856 G; 607 T; 0 U; 0 Other;  
  
Alignment Scores:  
Pred. No.: 0 Length: 2993  
Score: 3726.50 Matches: 730  
Percent Similarity: 83.33% Conservative: 0  
Best Local Similarity: 83.33% Mismatches: 6  
Query Match: 96.29% Indels: 141  
DB: 4 Gaps: 1  
  
US-10-045-815-4 (1-736) x AAH16288 (1-2993)  
  
Qy 1 MetAlaValArgAlaLeuLysLeuLeuThrThrLeuLeuAlaValAlaAlaSer 20  
Db 42 ATGGCGGTACGGGGTGAAGCTGCTGACACACTGCTGGCTGCTGGCGCGCTGCC 101  
Qy 21 GlnAlaGluValGluSerGluAlaGlyTyrGlyMetValThrProAspLeuPheAla 40

Db 102 CAGGCCGAGGTGCGATCCGAGGAGGATGGGCGCATGGTGAGCGCTGATCTCTTCGCC 161  
Qy 41 GluGlyThrAlaAlaTyrAlaArgGlyAspTrpProGlyValValLeuSerMetGluArg 60  
Db 162 GAGGGGACCGAGCTACGCGCGGAGCTGGCCCGGGTGGTCTCGAGCATGGAACGG 221  
Qy 61 AlaLeuArgSerArgAlaAlaLeuArgAlaLeuArgCysArgThrGlnCysAla 80  
Db 222 GCGCTGCGCTCCCGGCGAGCCCTCGCGCCCTTCGCTGCGCTGCCGACCCAGTGTGCC 281  
Qy 81 AlaAspPheProTrpGluLeuAspProAspTrpSerProSerProAlaGlnAlaSerGly 100  
Db 282 GCCGACTTCCGCTGGAGCTGGACCCGACTGGTCCCGCCAGCCGCGCCAGCCCTCGGCG 341  
Qy 101 AlaGlyAlaLeuArgAspLeuSerPhePheGlyGlyLeuLeuArgArgAlaAlaCysLeu 120  
Db 342 GCGCGCGCCCTCGCGAGCTGAGCTTCTTCGCGGGGCGCTTCTGCGTGGCGCTGCCCTG 401  
Qy 121 ArgArgCysLeuGlyProProAlaAlaHisSerLeuSerGluGluMetGluLeuPhe 140  
Db 402 CGCGCTGCTCGCGCGCCCGCGCCGCGCCACCTCGCTCAGCGAAGAGATGGAGTTC 461  
Qy 141 ArgLysArgSerProTrpAsnTyrLeuGlnValAlaTyrPheLysIleAsnLysLeuGlu 160  
Db 462 CGCAAGCGGAGCGCCCTCAACACTTACCTGCGAGTGGCTACTTCAAGATCAACAAGTTGAG 521  
Qy 161 LysAlaValAlaAlaAlaHisThrPhePheValGlyAsnProGluHisMetGluMetGln 180  
Db 522 AAAGTGTGTGTCGAGCACACACCTTCTTGTGGGCAATCTCTGAGCACATGGAAATGCGAG 581  
Qy 181 GlnAsnLeuAspTyrTyrGlnThrMetSerGlyValLysGluAlaAspPheLysAspLeu 200  
Db 582 CAGAACCTAGACTATTATCCAAACCATGTCTGAGTGAAGGAGGCGGACTTCAAGGATCTT 641  
Qy 201 GluThrGlnProHisMetGlnLysPheArgLeuGlyValArgLeuTyrSerGluGluGln 220  
Db 642 GAGACTCAACCCCATATGCAAGANTTCGACTGGGAGTGGGACTCTACTTCAGAGAAACAG 701  
Qy 221 ProGlnGluAlaValProHisLeuGluAlaAlaLeuGlnGluTyrPheValAlaTyrGlu 240  
Db 702 CCACAGGAAGCTGTGCCCCACCTAGAGGGGCGCTGCAAGATACTTTGTGSCCTATGAG 761  
Qy 241 GluCysArgAlaLeuCysGluGlyProTyrAspTyrAspGlyTyrAsnTyrLeuGluTyr 260  
Db 762 GAGTGGCGTGGCTCTGCGAGGGCGCTATGACTACGATGGCTACAACTACCTTGGATAC 821  
Qy 261 AsnAlaAspLeuPheGlnAlaIleThrAspHisTyrIleGlnValLeuAsnCysLysGln 280  
Db 822 AACGCTGACCTCTCCAGGCCATCAGATCATTAATCCAGTCTCTCACTGTAAAGCAG 881  
Qy 281 AsnCysValThrGluLeuAlaSerHisProSerArgGluLysProPheGluAspPheLeu 300  
Db 882 AACTGTGTCAAGGAGCTTGTCTCCCAAGTCCGAGAGAGCCCTTTGAAGACTTCCTC 941  
Qy 301 ProSerHisTyrAsnTyrLeuGlnPheAlaTyrTyrAsnIleGlyAsnTyrThrGlnAla 320  
Db 942 CCATCGCATTAATAATTATCTGCAGTTGCTTACTATAACATTTGGGAATTTATACAGGCT 1001  
Qy 321 GlyGluCysAlaLysThrTyrLeuLeuPhePheProAsnAspGluValMetAsnGlnAsn 340  
Db 1002 GTTGAATGGCCAGACCTATCTTCTTCTTCCCAATGACAGGATGATGAACCAAT 1061  
Qy 341 LeuAlaTyrTyrAlaAlaValMetLeuGlyGluGluHisThrArgSerIleGlyProArgGlu 360  
Db 1062 TTGGCCATATTATGCGACTATGCTTGGAGAAGAACACACCCAGATCCATCGGCCCGCTGAG 1121  
Qy 361 SerAlaLysGluTyrArgGlnArgSerLeuLeuGluLysGluLeuPhePheAlaTyr 380  
Db 1122 AGTGCCAGAGAGTACCGACAGCGAAGCTTACTTGGAAAAGAAAGACTGCTTTCTTCCTTAT 1181  
Qy 381 AspValPheGlyIleProPheValAspProAspSerTrpThrProGluGluValIlePro 400

Db 1182 GATGTTTTTGGAAATCCCTTTGTGATCCGATTCATGGACTCCAGGAGAGTGAATCCC 1241  
Qy 401 LysArgLeuGlnGluGlySerGlnArgGluThrAlaValArgLleSerGlnGlu 420  
Db 1242 AAGAGATTGCAGAGAAACAGAGATGAGAACGGGAAACAGCGGTACGCATCTCCAGGAG 1301  
Qy 421 IleGlyAsnLeuMetLysGluIleGluThrLeuValGluGluLysThrLysGluSerLeu 440  
Db 1302 ATTGGGAACCTTATGAAGAAATCGAGACCCCTTGTGGAAGAGAGAACCAAGAGTCACTG 1361  
Qy 441 AspValSerArgLeuThrArgGluGlyGlyProLeuLeuThrGluGlyLleSerLeuThr 460  
Db 1362 GATGTAGCAGACTGACCGGGAAGGTGGCCCTCTGTATGAAGGCATCAGTCTCACC 1421  
Qy 461 MetAsnSerLysLeuLeuAsnGlyTyrglnArgValValMetAspGlyValLleSerAsp 480  
Db 1422 ATGAACTCCAACTCCTCAATGGTTCCAGCGGTGTGTGATGGAGCGGTAAATCTCTGAC 1481  
Qy 481 HisGluCysGlnGluLeuGlnArgLeuThrAsnValAlaIleThrSerGlyAspGlyTy 500  
Db 1482 CACGAGTGTCCAGAGCTGCAGAGACTGACCAATGTGGCAGCAACTCAGGAGATGGCTAC 1541  
Qy 501 ArgGlyGlnThrSerProHisThrProAsnGluLysPheTyrglyValThrValPheLys 520  
Db 1542 CGGGTCAAGACTCCCCACATCTCCCAATGAAAGTTCTATGGTGCATCTCTTCAA 1601  
Qy 521 AlaLeuLysLeuGlyGlnGluLysValProLeuGlnSerAlaHisLeuTyTyAsn 540  
Db 1602 GCCCTCAAGCTGGGCAAGAGCAAAAGTTCCTCTGCAGAGTGCCACCTGTACTACAAC 1661  
Qy 541 ValThrGluLysValArgArgLleMetGlnSerTyrglyPheArgLeuAspThrProLeuTy 560  
Db 1662 GTGACGGAGAGGTGGCGCATCATGGAGTCTTCCCTCCCTGGATACGCCCTCTAC 1721  
Qy 561 PheSerTySerHisLeuValCysArgThrAlaIleGluGluValGlnAlaGluArgLys 580  
Db 1722 TTTTCTTCTACTCTATCTGTGTGGCGACTCCATCGAGAGGTCCAGGAGAGGAGAG 1781  
Qy 581 AspAspSerHisProValHisValAspAsnCysLleLeuAsnAlaGluThrLeuValCys 600  
Db 1782 GATGATAGTATCCAGTCCAGTGCAGCAACTGCATCTCTGATGCGGAGACCTCGTGTGT 1841  
Qy 601 ValLysGluProProAlaTyThrPheArgAspTyrglySerAlaIleLeuTyLeuAsnGly 620  
Db 1842 GTCAAAGAGCCCCAGCCCTTCCGCACTACAGCGCCATCTTTTACCTTAATGGG 1901  
Qy 621 AspPheAspGlyGlyAsnPheTyrglyPheThrGluLeuAspAlaTyThrValThrAlaGlu 640  
Db 1902 GACTTCGATGCGGAAACTTTTATTTTCACTGAATGATGCCAAGACCGTGAAGCGGAG 1961  
Qy 641 ValGlnProGlnCysGlyArgAlaValGlyPheSerSerGlyThrGluAsnProHisGly 660  
Db 1962 GTGAGAGCTCAGTGTGGAAGAGCGGTGGGATCTCTTCAGGCATGAAACCCACATGGA 2021  
Qy 661 ValLysAlaValThrArgGlyGlnArgCysAlaIleAlaLeuTrpPheThrLeuAspPro 680  
Db 2022 GTGAAGGCTGTACACAGGGGCGAGCGGTGTGCCATCGCCCTGTGTGTTTCACTCCCTGGACCT 2081  
Qy 681 ArgHisSerGluArg----- 685  
Db 2082 CGACACAGCGAGCG- GGTGAGAGCAGCTCGAGCGGGTGAAGCAGCTGGTGTGTGTGTGA 2140  
Qy 685 ----- 685  
Db 2141 CCCGTTCCAGAGCGCCCTTGTTGGCTTTCTTCTTCTCCCAATCCATTCAGAGTGGCT 2200  
Qy 685 ----- 685  
Db 2201 GAGACAGAAAGAGACACTTGGGACACCAAGCTCCACGCCCTGTCTATTATGTGTCACATTG 2260  
Qy 685 ----- 685  
Db 2261 CCTTGTCTCCTCGGCTGTGTGTGAACGGGATCCAGGTGGGAAAGAGGTCAAGACAGG 2320

Qy 685 ----- 685  
Db 2321 GAGCGATGCTGAGTTCTTGGTTCCCTCTTGGGCCCCACTTCAGTGTCTTTCAGAG 2380  
Qy 685 ----- 685  
Db 2381 AGTAGGACCTGCTGGGAAGGAGATGAGCCTGGGGCCATTAAAGAACCTTCCTTGTCCCT 2440  
Qy 685 ----- 685  
Db 2441 GGAAGTAGCAGCTGAGAGATAGCAGTGTCTGGAGCGGAGCCCTCTCTGAATGGGCAGG 2500  
Qy 686 -----AspArgValGlnAlaAspAspLeuValLysMetLeuPheSerPro 700  
Db 2501 GGTTCCTCTTCAGGACAGGCGTGCAGGAGATGACCTGGTGAAGATGCTCTTCAGCCCA 2560  
Qy 701 GluGluMetAspLeuSerGlnGluGlnProLeuAspAlaGlnGlnGlyProProGluPro 720  
Db 2561 GAAGAGATGCTCTCTCCAGGAGCAGCCCTGGATGCCAGCAGGCGCCCGCCGAGCCT 2620  
Qy 721 AlaGlnGluSerLeuSerGlySerGluSerLysProLysAspGluLeu 736  
Db 2621 GCACAGAGTCTCTCTCAGGAGTGAATCGAAGCCCAAGAGTGAAGCTA 2668

## RESULT 13

ABZ11345  
ID ABZ11345 standard; cDNA; 2152 BP.  
XX  
AC ABZ11345;  
XX  
DT 20-JAN-2003 (first entry)  
XX  
DE Human polynucleotide SEQ ID NO 227.

XX Human; genome mapping; gene therapy; food supplement; virus; fungus;  
XX cell-proliferative disorder; neurodegenerative disease; bacterial;  
XX Parkinson's disease; Alzheimer's disease; autoimmune disease;  
XX multiple sclerosis; diabetes; genetic disorder; wound; burn; infection;  
XX arthritis; cytotoxic; immunomodulator; nootropic; neuroprotective;  
XX antiparkinsonian; antidiabetic; immunosuppressive; dermatological;  
XX haemostatic; vulnery; fungicide; antibacterial; virucide; protozoacide;  
XX antiarthritic; gene; ss.  
XX Homo sapiens.

XX WO200270539-A2.

XX 12-SEP-2002.

XX 05-MAR-2002; 2002WO-US005095.

XX 05-MAR-2001; 2001US-00799451.

XX (HYSE-) HYSEQ INC.

XX Tang YT, Zhou P, Goodrich RW, Asundi V, Zhang J, Zhao QA, Ren F;  
XX Xue AJ, Yang Y, Ma Y, Yamazaki V, Chen R, Wang Z, Ghosh M,  
XX Wehrman T, Wang J, Wang D, Drmanac RT;

XX WPI; 2002-759812/82.  
XX P-PSDB; ABP69128.

XX New polynucleotides comprising sequences assembled from expressed  
XX sequence tags (ESTs), useful for treating cell-proliferative, or platelet  
XX neurodegenerative, autoimmune, genetic, myeloid or lymphoid, or platelet  
XX or coagulation disorders.

XX Claim 1; SEQ ID NO 227; 1012pp + Sequence Listing; English.

XX The invention relates to an isolated polynucleotide (I) comprising a  
XX nucleotide sequence selected from any of 948 sequences (ABZ11119-  
XX ABZ12066) or their mature protein coding portion, active domain coding



protein or complementary sequences. The polynucleotides are useful for identifying expressed genes or for physical mapping of human genome. The encoded polypeptides (ABP68902-ABP68949) are useful as molecular weight markers, as a food supplement, for generating antibodies, in medical imaging, screening and diagnostic assays and for treating cell-proliferative disorders (cancer), neurodegenerative diseases (Parkinson's or Alzheimer's disease), autoimmune diseases (multiple sclerosis, diabetes, lupus) genetic disorders, myeloid or lymphoid disorders, platelet or coagulation disorders, wound, burns, incision, ulcers, liver or lung fibrosis, infections (bacterial, viral, fungal, parasitic), arthritis, etc. Note: The sequence data for this patent did not form part of the printed specification but was obtained in electronic format directly from WIPO at ftp.wipo.int/pub/published\_pct\_sequences

XX Sequence 2152 BP; 494 A; 636 C; 602 G; 420 T; 0 U; 0 Other;

Alignment Scores:

Pred. No.: 0 Length: 2152  
Score: 3645.00 Matches: 699  
Percent Similarity: 95.24% Conservatives: 2  
Best Local Similarity: 94.97% Mismatches: 5  
Query Match: 94.19% Indels: 30  
DB: 6 Gaps: 1

US-10-045-815-4 (1-736) x ABZ11345 (1-2152)

Qy 1 MethAlaValArgAlaLeuLysLeuThrThrLeuLeuAlaValAlaAlaSer 20  
Db 32 ATGGCGGTACCGCGGTGAAAGCTGCTGACACACTGCTGGTGTGGCGGTGCGCTCC 91  
Qy 21 GlnAlaGluValGluSerGluAlaGlyTyrGlyMetValThrProAspLeuLeuPheAla 40  
Db 92 CAAGCCGAGGTGCGAGTCCGAGGCGAGATGGGGCATGGTGCCTGATCTGCTCTTGGCC 151  
Qy 41 GluGlyThrAlaAlaTyrAlaArgGlyAspTyrProGlyValValLeuSerMetGluArg 60  
Db 152 GAGGGGACCGCAGCTACGCGCGGGGACTGGCGCGGGGTGTCTTCAGCATGGGAACGG 211  
Qy 61 AlaLeuArgSerArgAlaAlaLeuArgAlaLeuArgLeuArgCysArgThrGlnCysAla 80  
Db 212 GCGTGGCTCCCGGGGAGCCCTCCGCGCCCTTCGCTGCTGCTGCCACCGAGTGTGCC 271  
Qy 81 AlaAspPheProTyrGluLeuAspProAspTyrSerProSerProAlaGlnAlaSerGly 100  
Db 272 CCGGACTTCCCGTGGAGCTGGACCCGACCTGGTCCCGCCAGCGCGCGCGCTCGGGC 331  
Qy 101 AlaGlyAlaLeuArgAspLeuSerPhePheGlyGlyLeuLeuArgArgAlaAlaCysLeu 120  
Db 332 CCGCGCGCCCTGCGCGACCTGAGCTTCTTCGGGGGCGCTTCGCTGCGCTGCGCTGCTG 391  
Qy 121 ArgArgCysLeuGlyProProAlaAlaHisSerLeuSerGluGluMetGluLeuGluPhe 140  
Db 392 CGCGCTGCTCGGGCGCGCGCGCCGACCTCGCTCAGCGAAGAGATGGAGTGGAGTTC 451  
Qy 141 ArgLysArgSerProTyrAsnTyrLeuGlnValAlaTyrPheLysIleAsnLysLeuGlu 160  
Db 452 CGCAAGCGGAGCCCTACAACTACTCGAGGTCGCTACTTCAAGATCAACAAGTGGAG 511  
Qy 161 LysAlaValAlaAlaAlaHisThrPhePheValGlyAsnProGluHisMetGluMetGln 180  
Db 512 AAAGCTGTGTGAGGACACACACTTCTTCGGGGCAATCTTCGAGCATGGAAATGCAG 571  
Qy 181 GlnAsnLeuAspTyrTyrGlnThrMetSerGlyValLysGluAlaAspPheLysLeu 200  
Db 572 CAGAACCTAGACTATTACAAACCATGTCGGAGTGAAGGAGGCGGACTTCAAGGATCTT 631  
Qy 201 GluThrGlnProHisMetGlnGluPheArgLeuGlyValArgLeuTyrSerGluGluGln 220  
Db 632 GAGACTCAACCCCATATCAAGAAATTTCGACTGGGAGTGGGACTCTACTCAGAGGAACAG 691  
Qy 221 ProGlnGluAlaValProHisLeuGluAlaAlaLeuGlnGluTyrPheValAlaTyrGlu 240  
Db 692 CCACAGGAGCTGTGCCCCCCTAGAGCGCGCTGCAAGAATACTTTGTGGCCTATGAG 751

Qy 241 GluCysArgAlaLeuCysGluGlyProTyrAspTyrAspGlyTyrAsnTyrLeuGluTyr 260  
Db 752 GAGTGGCGTGGCTCTGCGAAGGGCCCTATGACTAGTATGGTACAACTACCTTGGTAC 811  
Qy 261 AsnAlaAspLeuPheGlnAlaIleThrAspHisTyrIleGlnValLeuAsnCysLysGln 280  
Db 812 AACGCTGACCTTTCGAGCCATACAGATCATATCATCCAGTCTCACTGTAGGAG 871  
Qy 281 AsnCysValThrGluLeuAlaSerHisProSerArgGluLysProPheGluAspPheLeu 300  
Db 872 AACTGTGTCAAGGAGCTTGTCTCCCAAGTGCAGAGAAGCCCTTTGAAGACTTCCTC 931  
Qy 301 ProSerHisTyrAsnTyrLeuGlnPheAlaTyrTyrAsnIleGlyAsnTyrThrGlnAla 320  
Db 932 CCATCGCATTAATAATATCTGCAGTTTGCCTACTATAAATTTGGGAATATACAGGCT 991  
Qy 321 GlyGluCysAlaLysThrTyrLeuLeuPhePheProAsnAspGluValMetAsnGlnAsn 340  
Db 992 GTTGAATGGCCAGACCTATCTCTCTTCTCCCAATGACGAGTGTATGAACCAAAAT 1051  
Qy 341 LeuAlaTyrTyrAlaAlaMetLeuGlyGluGluHisThrArgSerIleGlyProArgGlu 360  
Db 1052 TTGGCCTATTATGCAGCTATGCTTGGAGAAGAACACACAGATCCATCGGCCCTGAG 1111  
Qy 361 SerAlaLysGluTyrArgGlnArgSerLeuLeuGluLysGluLeuPhePheAlaTyr 380  
Db 1111 ----- 1111  
Qy 381 AspValPheGlyIleProPheValAspProAspSerTyrThrProGluGluValIlePro 400  
Db 1112 -----GATTCTATGGACTCCAGAAAGAGTATTCC 1141  
Qy 401 LysArgLeuGlnGluLysGlnLysSerGluArgGluThrAlaValArgIleSerGlnGlu 420  
Db 1142 AAGAGATTCCAGAGAAACAGAAAGTCCAGAACCGGAAACAGCCGTACGCATCTCCACGAG 1201  
Qy 421 IleGlyAsnLeuMetLysGluIleGluThrLeuValGluLysThrLysGluSerLeu 440  
Db 1202 ATTGGAACTTATGAAGCAATCGAGACCTTGTGGAGAAGAACCAAGAGTCACTG 1261  
Qy 441 AspValSerArgLeuThrArgGluGlyProLeuLeuTyrGluGlyIleSerLeuThr 460  
Db 1262 GATGTAGCAGACTGACCCGGGAAGTGGCCCTCTGTGTATGAAGGCATATAGTCTCAC 1321  
Qy 461 MetAsnSerLysLeuLeuAsnGlyTyrGlnArgValValMetAspGlyValIleSerAsp 480  
Db 1322 ATGAATCCAACTCTTGTATGTTCCAGCGGTGTGTATGGCGGCGTAATCTCTGAC 1381  
Qy 481 HisGluCysGlnGluLeuGlnArgLeuThrAsnValAlaAlaThrSerGlyAspGlyTyr 500  
Db 1382 CACCAATGTCAAGAGCTGCAGAGACTGACCAATGGGGCAGCAACCTCAGAGATGGCTAC 1441  
Qy 501 ArgGlyGlnThrSerProHisThrProAsnGluLysPheTyrGlyValThrValPheLys 520  
Db 1442 CGGGGTGAGACTCCCCACATCTCCCAATGAAGAAGTTCATGGTGTCTACTGTTCAAA 1501  
Qy 521 AlaLeuLysLeuGlyGlnGluLysValProLeuGlnSerAlaHisLeuTyrTyrAsn 540  
Db 1502 GCCTCAAGCTGGGGCAAGAGCAAAAGTTCCTCTGCAGAGTGGCCACCTGTACTACAC 1561  
Qy 541 ValThrGluLysValArgArgIleMetGluSerTyrPheArgLeuAspThrProLeuTyr 560  
Db 1562 GTGACGGAGAAAGTGGCGGCATCATGGAGTCTCTACTCCGCTGGATAGCCCTCTTAC 1621  
Qy 561 PheSerTyrSerHisLeuValCysArgThrAlaIleGluGluValGlnAlaGluArgLys 580  
Db 1622 TTTTCTACTCTCATCTGGTGTGGCCACTGCCACTGCAAGAGTCCAGCAGAGAGAG 1681  
Qy 581 AspAspSerHisProValHisValAspAsnCysIleLeuAsnAlaGluThrLeuValCys 600  
Db 1682 GATGATAGTCACTCCAGTCCACGCTGGCAACTGCATCTCTGAATGCCGAGACCTCGTGT 1741

QY 601 VallysGluProAlaValThrPheArgAspTyrSerAlaIleLeuTyrLeuAsnGly 620  
 DB 1742 GTCARAGAGCCCGCCAGCTACACCTTCGCGACTACAGCGCCATCTTACCTAATGGG 1801  
 QY 621 AspPheAspGlyGlyAspPheTyrPheThrGluLeuAspAlaLysThrValThrAlaGlu 640  
 DB 1802 GACTTCGATGGCGGAACTTTTATTTCACTGAACCTGGATGCCAAGACCGTGCAGCGCAGAG 1861  
 QY 641 ValGlnProGlnCysGlyArgAlaValGlyPheSerSerGlyThrGluAsnProHisGly 660  
 DB 1862 GTGCGAGCTCAGTGTGGAAGAGCCGTGGGATCTCTTCAGGCATGAAAACCCACATGGA 1921  
 QY 661 VallysAlaValThrArgGlyGlnArgCysAlaIleAlaLeuTyrPheThrLeuAspPro 680  
 DB 1922 GTGAAGGCTGTCAACAGGGGCGAGCGCTGTGCCATCGCCCTGTGGTTACCCCTGGACCCCT 1981  
 QY 681 ArgHisSerGluArgAspArgValGlnAlaAspPheValLysMetLeuPheSerPro 700  
 DB 1982 CGACACGCGGCGGAGCAGGGTGCAGGCAGATGACCTGTGTGAAGATGCTCTTCAGGCCCA 2041  
 QY 701 GluGlnMetAspLeuSerGlnGluGlnProLeuAspAlaGlnGlnGlyProGluPro 720  
 DB 2042 GAAGAGATGGACCTCTCCAGAGCAGCGCCCTGGATGCCAGCAGGGGCCCGGACCT 2101  
 QY 721 AlaGlnGluSerLeuSerGlySerGluSerLysProLysAspGluLeu 736  
 DB 2102 GCACAAGAGTCTCTCTCAGGCGATGAATCGAAGCCCAAGGATGAGCTA 2149

## RESULT 14

AA151752  
 ID AA151752 standard; cDNA; 2127 BP.

AC AA151752;

DT 26-FEB-2002 (first entry)

DE DNA encoding novel secreted protein #1.

XX Secreted protein; cytostatic; immunosuppressive; vulnary; vaccine;  
 KW antiinflammatory; neuroprotective; nephrotropic; cardiovascular; human;  
 KW cancer; autoimmune disease; wound healing disorder; infection;  
 KW haematopoietic disorder; inflammatory disorder; infertility;  
 KW neurological disease; psychiatric disease; cardiovascular disease;  
 KW respiratory disease; renal; gastrointestinal; ss.

OS Homo sapiens.

XX Key Location/Qualifiers

FT 1. .2127

FT CDS /tag= a

FT /product= "Human secreted protein"

XX WO200179454-A1.

XX 25-OCT-2001.

XX 11-APR-2001; 2001WO-US011797.

XX 13-APR-2000; 2000US-0196603P.

XX 24-APR-2000; 2000US-0199417P.

XX (SMIK ) SMITHKLINE BEECHAM CORP.

XX (SMIK ) SMITHKLINE BEECHAM PLC.

XX Agarwal P, Murdoch PR, Rizvi SK, Smith RF, Xiang Z;

XX WPI; 2002-061975/08.

XX P-PSDB; AAU09860.

XX New secreted proteins or polypeptides, useful for treating e.g. cancer,  
 PT autoimmune diseases, wound healing disorder, infections, hematopoietic  
 PT disorders, inflammatory disorders, infertility, cancer.

PS Claim 2; Page 34-35; 92pp; English.

XX The invention relates to an isolated novel secreted polypeptide (I) and  
 CC polynucleotide (II). (I) and (II) are useful for treating cancer,  
 CC autoimmune diseases, wound healing disorder, infections, hematopoietic  
 CC disorders, inflammatory disorders, infertility, neurological and  
 CC psychiatric diseases, cardiovascular diseases, respiratory diseases,  
 CC renal diseases, or gastrointestinal diseases. These may also be used to  
 CC treat diseases, abnormalities and disorders caused by abnormal  
 CC expression, production, function and/or metabolism of the genes, as  
 CC vaccines for inducing immunological response in a mammal, and in  
 CC screening methods for detecting the effect of added compounds on the  
 CC production of mRNA and polypeptide in cells. The polypeptides can be used  
 CC as immunogens to produce antibodies immunospecific for the polypeptides,  
 CC and to identify membrane-bound or soluble receptors. The polynucleotides  
 CC may be used as diagnostic reagents, in chromosome localisation studies,  
 CC and in tissue expression studies. The present sequence represents the  
 CC coding sequence of novel human secreted protein #1

XX Sequence 2127 BP; 495 A; 623 C; 589 G; 420 T; 0 U; 0 Other;

## Alignment Scores:

Prod. No.: 0 Length: 2127  
 Score: 3572.00 Matches: 691  
 Percent Similarity: 94.57% Conservative: 5  
 Best Local Similarity: 93.89% Mismatches: 12  
 Query Match: 92.30% Indels: 28  
 DB: 6 Gaps: 3

US-10-045-815-4 (1-736) x AA151752 (1-2127)

QY 1 MetAlaValArgAlaLeuLysLeuLeuThrThrLeuLeuAlaValAlaAlaLaser 20  
 DB 1 ATGGGGGTACGGCGGTGAAGCTGCTGACACACACTGTGGTGTGTCGGCGCTGCCCTCC 60  
 QY 21 GlnAlaGluValGluSerGluAlaGlyTyrGlyMetValThrProAspLeuPheAla 40  
 DB 61 CAAGCGAGGTGCGAGTCCGAGCAGATGGGCGATGGTACGCTGATCTGCTCTTCGCC 120  
 QY 41 GluGlyThrAlaAlaTyrAlaArgGlyAspTyrProGlyValValLeuSerMetGluArg 60  
 DB 121 GAGGGGACCGCAGCCTACGCGCGCGGGGACTGGCCCGGGTGGTCTCTGAGCATGGAACGG 180  
 QY 61 AlaLeuArgSerArgAlaAlaLeuArgAlaLeuArgLeuArgCysArgThrGlnCysAla 80  
 DB 181 GCGTGGCTCCCGGGCAGCCCTCCGCGCCCTTCGCTGCGTCCGCCACCGAGTGGCC 240  
 QY 81 AlaAspPheProTyrGluLeuAspProAspTyrSerProSerProAlaGlnAlaSerGly 100  
 DB 241 GCGGACTTCCCGTGGGAGCTGGACCCCGACTGGTCCCCCAGCGCGCCAGGCTCGGGC 300  
 QY 101 AlaGlyAlaLeuArgAspLeuSerPhePheGlyGlyLeuLeuArgArgAlaAlaCysLeu 120  
 DB 301 GCGCGCGCCCTGCGGAGCCTGAGCTTCTCGGGGGCCCTTCGCGTGGCGCTGCCCTCG 360  
 QY 121 ArgArgCysLeuGlyProProAlaAlaHisSerLeuSerGluGluMetGluLeuPhe 140  
 DB 361 CGCGCTGCTCCCGCGCGCGCGCGCCCGCTCGCTCAGCGAAGAGATGGAGCTGGAGTTC 420  
 QY 141 ArgLysArgSerProTyrAsnTyrIleuGlnValAlaTyrPheLysIleAsnLysLeuGlu 160  
 DB 421 CGCAAGCGGAGCCCTACAACTACCTGAGTCCGCTACTTCAAGATCAACAAGTTGAG 480  
 QY 161 LysAlaValAlaAlaAlaHisThrPheValGlyAsnProGluHisMetGluMetGln 180  
 DB 481 AAAGCTGTGTGCGAGCACACACCTTCTTCGTGGCAATCTTGAGCACATCGAATGCAG 540  
 QY 181 GlnAsnLeuAspTyrTyrGlnThrMetSerGlyValLysGluAlaAspPheLysAspLeu 200  
 DB 541 CAGAACCTAGACTATTACCAACCATGTCTCGAGTGAAGGAGGCGGACCTTCAAGATCTT 600  
 QY 201 GluThrGlnProHisMetGlnGluPheArgLeuGlyValArgLeuTyrSerGluGln 220



Db 601 GAGACTCAACCCCATATGCAAGAAATTTGCGAGTGGGAGTGGACTTACTCAGAGGAACAG 660  
Qy 221 ProGlnGluAlaValProHisLeuGluAlaAlaLeuGlnGluTyrPheValAlaTyrGlu 240  
Db 661 CCACAGGAAGCTGTGCCCCACCTAGAGGGGGCTCAAGAAATACATTTTGTGGCTATGAG 720  
Qy 241 GluCysArgAlaLeuGluGluGluGluGluGluGluGluGluGluGluGluGluGluGlu 260  
Db 721 GAGTGGCGTGGCTCTGCGAAGGGGCTATGACTACGATGGCTACAACTACCTTGAGTAC 780  
Qy 261 AsnAlaAspLeuPheGlnAlaIleThrAspHisTyrIleGlnValLeuAsnGlnGlnGln 280  
Db 781 AACGCTGACCTCTTCAGGCGCATCAGATCATTATCCAGGTCTCAACTGTAAAGCAG 840  
Qy 281 AsnCysValThrGluLeuAlaSerHisProSerArgGluGluGluGluGluGluGluGlu 300  
Db 841 AACTGTGTCCAGGAGCTGTCTTCCCAAGTCGAGAGAGCCCTTTGAAGACTTCCTC 900  
Qy 301 ProSerHisTyrAsnTyrLeuGlnGlnPheAlaTyrTyrAsnIleGlyAsnTyrThrGlnAla 320  
Db 901 CCATCGATTAAATATATCTGCGAGTTTGGCTACTATAAC----- 939  
Qy 321 GlyGluCysAlaIysThrTyrLeuLeuPhePheProAsnAspGluValMetAsnGlnAsn 340  
Db 940 -----AAGACA-----ATCTGCTATTGTAAT 960  
Qy 341 LeuAlaTyrTyrAlaAlaMetLeuGluGluGluGluGluGluGluGluGluGluGluGlu 360  
Db 961 CTTCCCTGTCTCTCGAAATCTATAGAAA-----AAGAAG 996  
Qy 361 SerAlaIysGluTyrArgGlnArgSerLeuLeuGluGluGluGluGluGluGluGluGlu 380  
Db 997 AGTGCCNAGAGTACCGACAGCGAGCTACTCGAAAAGAACTGCTTTCTTCGCTTAT 1056  
Qy 381 AspValPheGlyIleProPheValAspProAspSerTyrThrProGluGluValIlePro 400  
Db 1057 GATGTTTTTGGAAATTCCTTTGTGGATCCGATTTCATGGACTCCAGAGAAGAGTATTCCT 1116  
Qy 401 LysArgLeuGlnGluGluGlnGluGluGluGluGluGluGluGluGluGluGluGluGlu 420  
Db 1117 AAGAGATTGCAAGAGAAACAGAGTACAGACGGGAAACAGCCGTACGCAATCTCCAGGAG 1176  
Qy 421 IleGlyAsnLeuMetLysGluIleGluThrLeuValGluGluGluGluGluGluGluGlu 440  
Db 1177 ATTGGGAACCTTATGAGGAATTCGAGACCTTTGTGGAGAGAAGACCAAGAGTCACTG 1236  
Qy 441 AspValSerArgLeuThrArgGluGluGluGluGluGluGluGluGluGluGluGluGlu 460  
Db 1237 GATGTGAGCAGACTGACCCCGGAGAGTGGCCCTGCTGTATGAGGCAATCAGTCTCACC 1296  
Qy 461 MetAsnSerLysLeuLeuAsnGlyTyrGlnArgValValMetAspGlyValIleSerAsp 480  
Db 1297 ATGNACTCCAAACTCTCTGAATGGTTCCAGCGGGTGTGTATGACGGCGTAATCTCTGAC 1356  
Qy 481 HisGluCysGlnGluLeuGlnArgLeuThrAsnValAlaAlaThrSerGlyAspGlyTyr 500  
Db 1357 CACGAGTGTGAGGAGTGCAGAGACTCACCATTGTCAGACCACTCAGGAGATGCTAC 1416  
Qy 501 ArgGlyGlnThrSerProHisThrProAsnGluGluGluGluGluGluGluGluGluGlu 520  
Db 1417 CGGGGTGAGACCTCCCACTACTCCCAATGAAAGTTCTATGGTGTCACTGTCTTCAA 1476  
Qy 521 AlaLeuLysLeuGluGlnGluGluGluGluGluGluGluGluGluGluGluGluGluGlu 540  
Db 1477 GCCCTCAAGCTGGGGCAAGAGGCAAGTTCTCTGAGAGTGCACCACTGTACTACAAC 1536  
Qy 541 ValThrGluLysValArgArgIleMetGluSerTyrPheArgLeuAspThrProLeuTyr 560  
Db 1537 GTGACGAGAGAGGTGCGCGCATCATGAGTCTTACTTCCGCTGGATACGCCCTCTAC 1596  
Qy 561 PheSerTyrSerHisLeuValCysArgThrAlaIleGluGluGluGlnAlaGluArgLys 580  
Db 1597 TTTTCTTACTCTCATCTGTTGTCGGCACTGCCATCGAAGAGTCCAGGCGAGAGGAG 1656

## RESULT 15

AAC64726

ID AAC64726 standard; cDNA; 2416 BP.

XX AAC64726;

XX 27-FEB-2001 (first entry)

DE Mouse tumour suppressor Gros1-L encoding cDNA SEQ ID NO:5.

KW Tumour suppressor; Gros1-L; Gros1-S; cell proliferation; regulation;

KW cancer; cytostatic; gene therapy; ss.

XX Mus musculus.

XX WO200065047-A1.

XX 02-NOV-2000.

XX 26-APR-2000; 2000WO-JP002731.

XX 26-APR-1999; 99JP-00118806.

XX (CHUG-) CHUGAI RES INST MOLECULAR MEDICINE INC.

XX Wadhwa R, Sugihara T, Yoshida A;

XX WPI: 2000-687340/67.

XX P-PSDB; AAB36393.

XX Full-length tumor suppressor genes encoding Gros1-L, Gros1-S and mouse

XX PT homologs participating in regulation of cell proliferation, useful in

XX development of preventives and remedies of cancer.

XX Claim 1; Page 83-91; 114pp; Japanese.

XX The present sequence encodes the mouse tumour suppressor designated Gros1

XX -L. Gros1-L and Gros1-S have cytostatic activity and can be used in gene

XX therapy. Gros1-L and Gros1-S genes are useful in the development of drugs

XX used to treat and prevent cancer

XX	SQ	Sequence	2416 BP; 579 A; 671 C; 683 G; 482 T; 0 U; 1 Other;
		Alignment Scores:	
	Pred. No.:	1-56e-298	Length: 2416
	Score:	3323.00	Matches: 630
	Percent Similarity:	92.37%	Conservative: 36
	Best Local Similarity:	87.38%	Mismatches: 52
	Query Match:	85.87%	Indels: 4
	DB:	3	Gaps: 2
	US-10-045-815-4	(1-736) x AAC64726 (1-2416)	
QY	11	ThrLeuLeuAlaValValAlaAlaAlaSer---	GlnAlaGluValGluSerGluAlaGly 29
DB	40	ACGATGCTAGCGTCCGCCCGCCCGCTTACGGGTTGGCGC-GAGTCTGAGCCGGGA	98
QY	30	TrpGlyMetValThrProAspLeuLeuPheAlaGluGlyThrAlaAlaTyrAlaArgGly	49
DB	99	TGGAGCGTGGCGCCCTGACCTGCTTTACGCAGAGGGGACCGCGGCTACTCCGCGAG	158
QY	50	AspTrpProGlyValValLeuSerMetGluArgAlaLeuArgSerArgAlaAlaLeuArg	69
DB	159	GACTGGCCCGGGTGTCTGCAACATGGAGCGGGCTCTGCGCTCGCGCGCGCCCTCGCT	218
QY	70	AlaLeuArgLeuArgCysArgThrGlnCysAlaAlaAspPheProTTrpGluLeuAspPro	89
DB	219	GCCTCTGGCTGGCTGGCGCACGCTGTGCCACCGAACTCCGCTGGCGCACCGACTG	278
QY	90	AspTrpSerProSerPro-----AlaGlnAlaSerGlyAlaGlyAlaLeuArgAspLeu	107
DB	279	GATCTCGGTCCGACCCCGACCTGAGCCAGGACCCCGCGCGCGCCCTGCGACGACTG	338
QY	108	SerPhePheGlyClyLeuLeuArgArgAlaAlaCysLeuArgArgCysLeuGlyProPro	127
DB	339	CGCTTCTTCGGAGCCGTGTGCGCGGTGCGCGCTGCCCTACCGCGTGCCTCGGCGCCGCC	398
QY	128	AlaAlaHisSerLeuSerGluGluMetGluLeuGluPheArgLysArgSerProTyrAsn	147
DB	399	TCTGCCCACTTGTGAGTGAAGACTGGACCTGGAGTTCAACAGCGGAGCGCCGTACAAC	458
QY	148	TyrLeuGlnValAlaTyrPheLysIleAsnLysLeuGluLysAlaValAlaAlaHis	167
DB	459	TACTTCAGGTGCGCTATTTCAGAGATAAACAGCTGGAGAAAGCTGTGGCTGGCGGCAC	518
QY	168	ThrPhePheValGlyValAsnProGluHisMetGluMetGlnGlnAsnLeuLeuAspTyrTyrGln	187
DB	519	ACCTTTCTTTGGGCGATCTCTGAGGCACATGGAGATGGCGGAGAACCTCGACTATTACAA	578
QY	188	ThrMetSerGlyValLysGluAlaAspPheLysAspLeuGluThrGlnProHisMetGln	207
DB	579	ACCATGCTGGGTGAAGAGGAGCAGACTTCAGGGATCTCGAGGCGCAAGCCCATATGCAT	638
QY	208	GluPheArgLeuGlyValArgLeuTyrSerGluGluGlnProGlnGluAlaValProHis	227
DB	639	GAGTTTCGGCTGGGGTACGACTCTACTCAGAGGAGAACCCACAGGAAGCTGTGCCCCAC	698
QY	228	LeuGluAlaAlaLeuGlnGluTyrPheValAlaTyrGluGluCysArgAlaLeuCysGlu	247
DB	699	CTGGAGCGGCACCTCAAGAGTACTTTGTGGCCGATGAGGAGTGCCGTGCCCTCTGCGAA	758
QY	248	GlyProTyrAspTyrAspClyTyrAsnTyrLeuGluTyrAsnAlaAspLeuPheGlnAla	267
DB	759	GGGCCCTATGACTACAGCGCTACAACTACCTAGACTACAGCGCTGACCTCTTCCAGGCC	818
QY	268	IleThrAspHisTyrIleGlnValLeuAsnCysLysGlnAsrCysValThrGluLeuAla	287
DB	819	ATCACAGATCATTACGTCCAGCTCTCAACTGTGAAGCAGAACCTGTGTACGGAGCTGGCT	878
QY	288	SerHisProSerArgGluLysProPheGluAspPheLeuProSerHisTyrAsnTyrLeu	307
DB	879	TCCCAACCAAGTAGGGAAGCCCTTTGAAGACTTCTCCCTCCCTACACTATAATTACCTA	938

Db	2019		CAGCGTGGCCATCGCCCTGTGTTTCACGCTGGATCCTCGGCACAGTGAGAGACAGG	2078
Qy	688		valGlnAlaAspLeuValLysMetLeuPheSerProGluGluMetAspLeuSerGln	707
Db	2079		GTGCAGGCAGATGACCTGGTGAAGATGCTGTTTCAGCCCAAGAGAGGTGGACCTCCCCCAG	2138
Qy	708		GluGlnProLeuAspAlaGlnGlnGlyProProGluProAlaGlnGluSerLeuSerGly	727
Db	2139		GAACAGCCCTGCTGACCAGCAGGGTTCGCCAGAGCCTGGAGAGAGTTTCTGCATGCT	2198
Qy	728	Ser	728	
Db	2199	GCT	2201	

Search completed: July 18, 2004, 10:20:02  
Job time : 1086 secs

This Page Blank (uspto)

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: July 18, 2004, 08:21:37 ; Search time 53 Seconds  
(without alignments)

723.087 Million cell updates/sec

Title: US-10-045-815-4

Perfect score: 3870

Sequence: 1 MAVRAKLKLTLLAVVAAS.....PPEPAQESLSGSEKPKDEL 736

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 141681 seqs, 52070155 residues

Total number of hits satisfying chosen parameters: 141681

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt\_42:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	572	14.8	401	1	CASP HUMAN
2	536.5	13.9	400	1	CASP MOUSE
3	501	12.9	437	1	NO55 HUMAN
4	500.5	12.9	431	1	SC65 RAT
5	419	10.8	271	1	CASP CHICK
6	157.5	4.1	603	1	UVR_C_HLCPN
7	124	3.2	1129	1	EG2T_CABEL
8	120.5	3.1	2291	1	SPCB_DROME
9	118.5	3.1	609	1	CEP2_MOUSE
10	118	3.0	496	1	MOQ3_STAEP
11	118	3.0	1597	1	CTRO_MOUSE
12	117	3.0	1286	1	CTRO_HUMAN
13	116	3.0	884	1	RASO_SULSO
14	116	3.0	1186	1	ADDB_BACSU
15	115.5	3.0	1969	1	MYSA_CABEL
16	114.5	3.0	613	1	SG2_BOVIN
17	114	2.9	928	1	NIBA_HUMAN
18	114	2.9	3911	1	AKA9_HUMAN
19	113	2.9	1127	1	H855_TREPA
20	113	2.9	1805	1	HMW2_MVCGE
21	112.5	2.9	1726	1	FBPA_AQUAE
22	112.5	2.9	1893	1	CSP2_HUMAN
23	112	2.9	887	1	NED4_MOUSE
24	112	2.9	1694	1	CLH_DICDI
25	111	2.9	2431	1	POLN_SFV
26	110.5	2.9	879	1	RA50_SULTO
27	110.5	2.9	1257	1	CCAA_BACTU
28	110	2.8	388	1	CUS5_HUMAN
29	110	2.8	859	1	MUTS_AQUAE
30	109.5	2.8	543	1	IEFS_HUMAN
31	109	2.8	1000	1	NED4_HUMAN
32	109	2.8	2230	1	GOA4_HUMAN
33	108.5	2.8	887	1	NED4_RAT

34 108 2.8 1084 1 MYSS\_RABIT  
35 108 2.8 2663 1 CENE\_HUMAN  
36 107.5 2.8 624 1 PRPD\_BACSU  
37 107.5 2.8 714 1 FRDA\_HELPY  
38 107 2.8 605 1 UVR\_C\_HLCPV  
39 106.5 2.8 349 1 CUS5\_MOUSE  
40 106 2.7 503 1 EGLX\_MOUSE  
41 106 2.7 705 1 YW2\_YEAST  
42 106 2.7 1109 1 YWSD\_DICDI  
43 105 2.7 693 1 EX70\_DROME  
44 105 2.7 3210 1 CENF\_HUMAN  
45 105 2.7 4687 1 PLEI\_RAT

#### ALIGNMENTS

##### RESULT 1

CASP\_HUMAN STANDARD; PRT; 401 AA.  
AC 075718;  
DT 16-OCT-2001 (Rel. 40, Created)  
DT 16-OCT-2001 (Rel. 40, Last sequence update)  
DT 10-OCT-2003 (Rel. 42, Last annotation update)  
DE Cartilage-associated protein precursor.  
GN CRTAP OR CASP.  
OS Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
OX NCBI\_TaxID=9606;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC TISSUE=Fetal brain;  
RX MEDLINE=20169181; PubMed=10702664;  
RA Tonachini L., Morello R., Monticone M., Skaug J., Scherer S.W.,  
RA Cancedda R., Castagnola P.,  
RT "cDNA cloning, characterization and chromosome mapping of the gene  
encoding human cartilage associated protein (CRTAP).";  
RL Cytogenet. Cell Genet. 87:191-194(1999).  
RN [2]

SEQUENCE FROM N.A.  
RC TISSUE=Kidney;  
RX MEDLINE=22388257; PubMed=12477932;  
RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,  
RA Klausner R.D., Collins P.S., Wagner L., Shenmen C.M., Schuler G.D.,  
RA Altschul S.F., Buetow K.H., Buetow K.H., Schaefer C.F., Bhat N.K.,  
RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,  
RA Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,  
RA Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,  
RA Brownstein M.J., Usdin T.B., Toshiyuki S., Carninci P., Prange C.,  
RA Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullaly S.J.,  
RA Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,  
RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,  
RA Villalón D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,  
RA Fahey J., Helton E., Kettner M., Maman A., Rodriguez S., Sanchez A.,  
RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,  
RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,  
RA Rodriguez A.C., Grumman J., Schmutz J., Myers R.M.,  
RA Butterfield Y.S.N., Krzywinski M.I., Skalska U., Smailus D.E.,  
RA Schnerch A., Schein J.E., Jones S.J.M., Marra M.A.,  
RT "Generation and initial analysis of more than 15,000 full-length  
human and mouse cDNA sequences.";  
RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).  
CC -!- SUBCELLULAR LOCATION: Secreted. Extracellular matrix (By  
similarity).  
CC -!- TISSUE SPECIFICITY: Found in articular chondrocytes. Expressed in  
a variety of tissues.  
CC -!- SIMILARITY: BELONGS TO THE CRTAP / NO55 FAMILY.

-----  
This SWISS-PROT entry is copyright. It is produced through a collaboration  
between the Swiss Institute of Bioinformatics and the EMBL outstation -  
the European Bioinformatics Institute. There are no restrictions on its  
use by non-profit institutions as long as its content is in no way

CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR ENBL; AJ006470; CAA07054.1; -  
 DR ENBL; BC008745; AA008745.1; -  
 DR Genew; HGNC:2379; CRTAP.  
 DR MIM; 605497; -  
 DR InterPro; IPR008940; Prenyl trans.  
 DR Extracellular matrix; Signal.  
 KW Extracellular matrix; Signal.  
 FT SIGNAL 1 26 POTENTIAL.  
 FT CHAIN 27 401 CARTILAGE-ASSOCIATED PROTEIN.  
 FT CARBOHYD 87 87 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 363 363 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CONFLICT 137 137 E -> D (IN REF. 2).  
 SQ SEQUENCE 401 AA; 46561 MW; 4BEED4089195456F CRC64;  
 Query Match 14.8%; Score 572; DB 1; Length 401;  
 Best Local Similarity 34.4%; Pred. No. 9.5e-30;  
 Matches 145; Conservative 68; Mismatches 168; Indels 40; Gaps 11;  
 QY 4 RALKLLTLLAVAA--ASQAEVSEAGWGVTPDLL-----FAEGTAAYARGDMPGVVL 56  
 DB 6 RGAALLLCVACALRAGRAQYERYSFRPRDELMPLESAYRHADKYSGEHWAEVSG 65  
 QY 57 SMERALSRAALRALRLRCCTCAADFPWELDPWSPSPAQASGAGALRDLSPFGGLLR 116  
 DB 66 YLEISLRLHLLRDLSEAFCHRNCSA-----APQPEPAAGLASYPELRFGGLLR 115  
 QY 117 AALRRRC-LGPPAAHSL--SEEMELEFRKSPYNYLQVAYFKINKLEKAVAAAHTEFVGN 173  
 DB 116 AHCLRCCKQGLPAFRQSPSEVLADFORREPKYFLOFAFKANLPKAAAHTELLKH 175  
 QY 174 PEHMEMQNLDYQTMGSVKAEADFKDLETPHMQEFLGVRLYSEOPQEAHPHLEAALQ 233  
 DB 176 PDDEMMRNWAYKSLFGAEDY-IKDLETKSYSLFIRAVRYNGENWRTSITDMELALP 234  
 QY 234 EYFVAYECCALCGPYDYGNYLYNADLFOAITDHYIQLVNCQNCVTELASHPSRE 293  
 DB 235 DFFKAFVECLAACGSEIKDFK-----DFYLSIADHYVEVLECKIQEENL-TPVIGG 287  
 QY 294 KPPEDFLPSPHYNYLQPAYNYGNITQAGECAKTYLLFFPNDVNMQNLYAYAM-----L 348  
 DB 288 YPVEKVFATWYHYLQFAYKLNDAKNAAPCAVSYLLFDQNDKVMQNLVYVYHRDTWGL 347  
 QY 349 GEETHRISIGRESAKYQRSLLEKELLFFAYDVFGIPFFVDPSWTPPEVPIKRLQEKOK 408  
 DB 348 SDEHFQ---PRPEAVQFNVTTLQKELYDPAKE-----NIMDDGEVVEYVDLLELET 400  
 QY 409 S 409  
 DB 401 S 401  
 RESULT 2  
 CASP MOUSE STANDARD; PRT; 400 AA.  
 ID -CASP MOUSE  
 AC Q9CYD3; O88698;  
 DT 16-OCT-2001 (Rel. 40, Created)  
 DT 16-OCT-2001 (Rel. 40, Last sequence update)  
 DT 10-OCT-2003 (Rel. 42, Last annotation update)  
 DE Cartilage-associated protein precursor.  
 GN CRTAP OR CASP.  
 OS Mus musculus (Mouse).  
 OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 OC NCBI\_TaxID=10090;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=Embryo;  
 RX MEDLINE=93357019; PubMed=10429950;  
 RA Morello R., Tonachini L., Monticone M., Viggiano L., Rocchi M.,  
 RA Cancedda R., Castagnola P.;

RT "cDNA cloning, characterization and chromosome mapping of Crtap  
 RT encoding the mouse cartilage associated protein.";  
 RN Matrix Biol. 18:319-324(1999).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=C57BL/6J; TISSUE=Embryo;  
 RX MEDLINE=21085660; PubMed=11217851;  
 RA Kawai J., Shinagawa A., Shibata K., Yoshino M., Itoh M., Ishii Y.,  
 RA Arakawa T., Hara A., Fukunishi Y., Konno H., Adachi J., Fukuda S.,  
 RA Aizawa K., Izawa M., Nishi K., Kiyosawa H., Kondo S., Yamana K.,  
 RA Saio T., Okazaki Y., Gojobori T., Bono H., Kasukawa T., Saito R.,  
 RA Kadota K., Matsuda H.A., Ashburner M., Batalov S., Casavant T.,  
 RA Fleischmann W., Gaasterland T., Gissi C., King B., Kochiwa H.,  
 RA Kuehl P., Lewis S., Matsuo Y., Nikolaio I., Pesole G., Quackenbush J.,  
 RA Schriml L.M., Staubli F., Suzuki R., Tomita M., Wagner L., Washio T.,  
 RA Sakai K., Okido T., Furuno M., Aono H., Baldarelli R., Barsh G.,  
 RA Blake J., Boffelli D., Bojunga N., Carninci P., de Bonaldo M.F.,  
 RA Brownstein M.J., Bult C., Fletcher C., Fujita M., Gariboldi M.,  
 RA Gustincich S., Hill D., Hofmann M., Hume D.A., Kamiya M., Lee N.H.,  
 RA Lyons P., Marchionni L., Mashima J., Mazzarelli J., Mombaerts P.,  
 RA Nordone P., Ring B., Ringwald M., Rodriguez I., Sakamoto N.,  
 RA Sasaki H., Sato K., Schoenbach C., Seya T., Shibata Y., Storch K.-F.,  
 RA Suzuki H., Toyooka K., Wang K.H., Weitz C., Whittaker C., Wilming L.,  
 RA Wynshaw-Boris A., Yoshida K., Hasegawa Y., Kawaji H., Kohsaki S.,  
 RA Hayashizaki Y.;  
 RT "Functional annotation of a full-length mouse cDNA collection.";  
 RL Nature 409:685-690(2001).  
 CC -!- SUBCELLULAR LOCATION: Secreted; extracellular matrix. Expressed in  
 CC -!- TISSUE SPECIFICITY: Found in articular chondrocytes. Expressed in  
 CC a variety of tissues.  
 CC -!- SIMILARITY: BELONGS TO THE CRTAP / NO55 FAMILY.  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR ENBL; AJ006469; CAA07053.1; -  
 DR ENBL; AK017797; BAB30938.1; -  
 DR MGD; MGI:1891221; Crtap.  
 DR InterPro; IPR008940; Prenyl trans.  
 KW Extracellular matrix; Signal.  
 FT SIGNAL 1 25 POTENTIAL.  
 FT CHAIN 26 400 CARTILAGE-ASSOCIATED PROTEIN.  
 FT CARBOHYD 86 86 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 362 362 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CONFLICT 7 7 A -> T (IN REF. 2).  
 FT CONFLICT 34 34 N -> S (IN REF. 2).  
 FT CONFLICT 79 80 SE -> RQ (IN REF. 2).  
 FT CONFLICT 249 250 SR -> VA (IN REF. 2).  
 FT CONFLICT 254 254 D -> T (IN REF. 2).  
 SQ SEQUENCE 400 AA; 46166 MW; 812245C6A4E769E2 CRC64;  
 Query Match 13.9%; Score 536.5; DB 1; Length 400;  
 Best Local Similarity 34.9%; Pred. No. 1.9e-27;  
 Matches 130; Conservative 64; Mismatches 145; Indels 33; Gaps 10;  
 QY 46 YARGDMPGVLSMERALRSRAALRALRLRCCTCAADFPWELDPWSPSPAQASGAGALR 105  
 DB 54 YSGEHWAEVSGYLEVSLRLHLLRDLSEAFCHRNCSAATP-----APAPA 103  
 QY 106 DLSPFGGLLRRAACLRRC-LGPPAAHSL--SEEMELEFRKSPYNYLQVAYFKINKLEKA 162  
 DB 104 ELRLFGSVLRRAQCLKRCQGLPAFRQSPSEVLADFORREPKYFLOFAFKANLPKA 163  
 QY 163 VAAAHTEFVGNPEHMEMQNLDYQTMGSVKAEADFKDLETPHMQEFLGVRLYSEOPQ 222  
 DB 164 IAAAHTEYLLRHPPDDMMKRNNEYKSLFPGA-EDHKLETKSYSLFRAVRYNGENWR 222  
 QY 223 EAVPHEAALQOEYFVAYECCALCGPYDYGNYLYNADLFOAITDHYIQLVNCQNC 282

Db 223 TSISDMELALPDFELKAFYECLACESREIKDFK-----DFYLSADHYVEVELEKIR 276  
Qy 283 VTSLASHPSREKPEFPLSHNYLOFAYYNYIOYQAGCAKTYLLFFPNDEVMNQNL 342  
Db 277 -BETLTPVIGYVPEKVFVATWYHLOFAYYKNDLKNAPCAVSYLLFFQSDRWMOQNLV 335  
Qy 343 YXAM-----LGEHRTSIPRESAKEYRQSLLEKELFFAFVDFVGFIPFVDPDSWTPEE 397  
Db 336 YQVHRDKWGLSDHFQ---PRPEAVQFNVTLQKELYDFQE-----HLMDDDEGEVVE 388  
Qy 398 VIPKRLQEKOKS 409  
Db 389 YVDDLLETESA 400  
RESULT 3  
ID NO55 HUMAN STANDARD; PRT; 437 AA.  
AC Q92791; Q9H476;  
DT 15-JUL-1998 (Rel. 36, Created)  
DT 15-JUL-1998 (Rel. 36, Last sequence update)  
DT 15-MAR-2004 (Rel. 43, Last annotation update)  
DE Nucleolar autoantigen NO55.  
GN SC65 OR NOL55.  
OS Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
OX NCBI\_TaxID=9606;  
RN [1]  
RP TISSUE=Bladder epithelium;  
RC MEDLINE=97015880; PubMed=8862517;  
RA Ochs R.L., Stein T.W. Jr., Chan E.K.L., Ruutu M., Tan E.M.;  
RT "cDNA cloning and characterization of a novel nucleolar protein.";  
RL Mol. Biol. Cell 7:1015-1024(1996).  
RN [2]  
RP SEQUENCE FROM N.A.  
RC TISSUE=Bladder epithelium;  
RX MEDLINE=2041121; PubMed=10952778;  
RA Fossa A., Siebert R., Aasheim H.C., Maelandsmo G.M., Berner A.,  
RA Fossa S.D., Paus E., Smeland E.B., Gaudernack G.;  
RT "Identification of nucleolar protein NO55 as a tumour-associated  
autoantigen in patients with prostate cancer.";  
RL Br. J. Cancer 83:743-749(2000).  
RN [3]  
RP SEQUENCE FROM N.A.  
RC TISSUE=Brain;  
RX MEDLINE=22388257; PubMed=12477932;  
RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,  
RA Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,  
RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,  
RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,  
RA Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,  
RA Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,  
RA Brownstein M.J., Ustin T.B., Toshiyuki S., Carninci P., Prange C.,  
RA Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullaly S.J.,  
RA Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,  
RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,  
RA Villalon D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,  
RA Fahy J., Helton E., Kettaman M., Madan A., Rodriguez S., Sanchez A.,  
RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,  
RA Blakeley R.W., Touchman J.W., Green E.D., Dickson M.C.,  
RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,  
RA Butterfield Y.S.N., Krzywinski M.I., Skalska U., Smalusz D.E.,  
RA Schnerch A., Schein J.E., Jones S.J.M., Marra M.A.;  
RT "Generation and initial analysis of more than 15,000 full-length  
human and mouse cDNA sequences.";  
RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).  
CC -I- SUBCELLULAR LOCATION: Nuclear; nucleolar. Localized uniformly  
throughout the granular component of the nucleolus and on the  
surface of chromosomes during mitosis.  
CC -I- SIMILARITY: BELONGS TO THE CRTAP / NOS5 FAMILY.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/  
CC or send an email to license@isb-sib.ch).  
CC -----  
CC EMBL; U47621; AAC51792.1; --  
CC EMBL; AJ250593; CAC16786.1; --  
CC EMBL; BC001047; AAH01047.1; --  
CC EMBL; BC007942; AAH07942.1; --  
CC EMBL; BC011701; AAH11701.1; --  
CC GK; Q92791; --  
CC GO; GO:000230; C:nuclear mitotic chromosome; TAS.  
CC GO; GO:0005730; C:nucleolus; TAS.  
CC GO; GO:0005716; C:synaptonemal complex; TAS.  
CC GO; GO:0007130; P:synaptonemal complex formation; TAS.  
CC InterPro; IPR008940; Prenyl\_trans.  
KW Nuclear protein; Antigen.  
FT DOMAIN 346 437 ASP/GLU-RICH (ACIDIC).  
FT CONFLICT 186 186 Q -> R (IN REF. 2).  
SQ SEQUENCE 437 AA; 50381 MW; 50C82FCE9BB7274A CRC64;  
Query Match 12.9%; Score 501; DB 1; Length 437;  
Best Local Similarity 33.3%; Pred. No. 4.1e-25;  
Matches 138; Conservative 53; Mismatches 191; Indels 32; Gaps 10;  
Qy 12 LLAVVAAASQAQVESEAGCGWVTPDLL-----PAEGTAAVARGDWPGVVLGSMERASRA 66  
Db 8 LLWLLGSAGAGYKYSFRGPPEDLPLAAAYGHAEQYEGESWRSEARVLEALRLHR 67  
Qy 67 ALRALRLRCRTQCAADFF-WELDPDWSFSPAQAAGALRDLSPFGGLLRRAACLRCLG 125  
Db 68 LLRDSEAFCHANGSGPAAPKPDGGRADWAC-----ELRLFGVRLERAACLRCKR 121  
Qy 126 PPAHSL---SEEMLEPRKSPNYLOVAVYKINKLEKAVAAAHFFVGHPHEMOMQON 182  
Db 122 TLPAQVPPYPPQLRLDFQSLPYQYGHYALFKANRLEKAVAAAYTFQRPKHELTAKY 181  
Qy 183 LDYYTNGSGVKEADFKLETPHMQEFLGVRLYSEQPAQVHLEALQEFVAYEEC 242  
Db 182 LNYQGLMDVADESITDLEAQPYEAVFLRAVKLYNSGDFRSTEDMERALSAYLAVPARC 241  
Qy 243 RALCEGPDYDGYNYLEYNADLFOAITDHYIQVNLKQNCVTELASHPSRKPFED-FLP 301  
Db 242 LAGCEGAHEQVDFK-----DFYPAIADLFAESLQCKYDCEANLT--PNVGGYFVDKFA 293  
Qy 302 SHYNYLOFAYYNYGNTOAGCAKTYLLFFPNDEVMNQNLAY--AAMLCEHTRSIGP 358  
Db 294 TMYHLOFAYYKNDVROAARSAASYMLDFDSDVMQONLYYFHRARWGLEE-EDFQP 352  
Qy 359 RESAKYRQSLLEKELFFAFVDFVGFIPFVDPDSWTPEEIPKRLQEKOKSERE 412  
Db 353 REEAMLYHNQTAELRELFETH---MYLQSDDEMELEETEPPELPEDALSDAE 402  
RESULT 4  
SC65 RAT  
ID SC65 RAT STANDARD; PRT; 431 AA.  
AC Q64375;  
DT 15-JUL-1998 (Rel. 36, Created)  
DT 15-JUL-1998 (Rel. 36, Last sequence update)  
DT 16-OCT-2001 (Rel. 40, Last annotation update)  
DE Synaptonemal complex protein SC65.  
GN SC65.  
OS Rattus norvegicus (Rat).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.  
OX NCBI\_TaxID=10116;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC TISSUE=Testis;

RX MEDLINE=92213429; PubMed=1363622;  
 RA Chen Q., Pearlman R.E., Moens P.B.;  
 RT "Isolation and characterization of a cDNA encoding a synaptonemal  
 RT complex protein."  
 RL Biochem. Cell Biol. 70:1030-1038(1992).  
 CC -!- SUBCELLULAR LOCATION: Nuclear. Located in the pairing zone of the  
 CC synaptonemal complex.  
 CC -!- TISSUE SPECIFICITY: Found in testis, brain, heart and at a much  
 CC lower level in liver.  
 CC -!- SIMILARITY: BELONGS TO THE CRTAP / NOS5 FAMILY.  
 CC  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC  
 CC EMBL; X65454; CAA46449.1; -.  
 DR PIR; A56822; A56822.  
 DR InterPro; IPI008940; Prenyl\_trans.  
 KW Nuclear protein.  
 FT DOMAIN 319 328 GLU-RICH (ACIDIC).  
 FT DOMAIN 353 384 GLU-RICH (ACIDIC).  
 SQ SEQUENCE 431 AA; 49995 MW; 4A34F3029407B2E7 CRC64;  
 Query Match 12.9%; Score 500.5; DB 1; Length 431;  
 Best Local Similarity 35.9%; Pred. No. 4.4e-25;  
 Matches 124; Conservative 47; Mismatches 153; Indels 21; Gaps 8;  
 Qy 46 YARDGFWGVLSMERALRSRAALRLRCRTCAADFPWELDPDWSPPS- --AQSAGAG 102  
 Db 14 YEGSWRESARYLEALRLHLLRLDSFAFCHANGCSG- PATSPRPAPGPDGNEGDDG 71  
 Qy 103 ALRDLSPFGGLLRRAACRLCLGPPAHSLS- --SEMELEFRKRSYNYLOVAFKINKL 159  
 Db 72 WARELRFLGHVLERAAALRLCKRTLPAPFPYPSRQLLRDFONRLPYQYLYAHAFKANRL 131  
 Qy 160 EKAVAAHTTFVGNPEHMEQNLIDYQTMGSKVEADFKDLETQPMQBEFLGLVRLYSEE 219  
 Db 132 EKAVAAATYFLQRPKELTAKYLYNYRGLMDIGDESILTDLEAQPYEAVFLQAVKLYNSG 191  
 Qy 220 QPQAVPHLEALQOYFVAYEBCALCEGPDYDGVNLYNADLFOAITDHYIQVNLCK 279  
 Db 192 DFRSSTHEMERALADYMTVFARCLAGCEGAHQVDFK- ----DFYPAIADUFAESLOCK 245  
 Qy 280 QNCVTELASHPREKPED- FLPSHYNYLOFAYNYIGNYTOAGECAKTYLLPFPDENVN 338  
 Db 246 VDCEANLT- -PNVGFFVDFKVAFTWYHYLOFAYKLVNDVQARSASAYMLFPKDSVMQ 303  
 Qy 339 QNLAYY- --AAMIGEHTRSIGPRESAKYQRQSLLEKELLFFAY 380  
 Db 304 QNLVYFRHRAWGLEE- EDFQPREAVLYHNTSELRLLELDFTH 347

## RESULT 5

ID -CASP\_CHICK STANDARD; PRT; 271 AA.  
 AC Q90830;  
 DT 15-JUL-1998 (Rel. 36, Created)  
 DT 15-JUL-1998 (Rel. 36, Last sequence update)  
 DT 10-OCT-2003 (Rel. 42, Last annotation update)  
 DE Cartilage associated protein precursor (Dualin).  
 GN CRTAP OR CASP.  
 OS Gallus Gallus (Chicken).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Archosauria; Aves; Neognathae; Galliformes; Phasianidae;  
 OC Gallus.  
 OX NCB1\_TaxID=9031;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=97360293; PubMed=9217321;

## RESULT 6

ID UVRC\_CHLPN STANDARD; PRT; 603 AA.  
 AC Q926W6; Q9JQAG;  
 DT 30-MAY-2000 (Rel. 39, Created)  
 DT 30-MAY-2000 (Rel. 39, Last sequence update)  
 DT 10-OCT-2003 (Rel. 42, Last annotation update)  
 DE UVrABC system protein C (Uvrc protein) (Excinuclease ABC subunit C).  
 GN UVRC OR CPN0940 OR CP0921 OR CP0974.  
 OS Chlamydia pneumoniae (Chlamydia pneumoniae).  
 OC Bacteria; Chlamydiae; Chlamydiales; Chlamydiaceae; Chlamydia.  
 OX NCB1\_TaxID=833558;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=CNL029;  
 RX MEDLINE=99206606; PubMed=10192388;  
 RA Kalman S., Mitchell W., Marathe R., Lammel C., Fan J., Hymen R.W.,  
 RA Olinger L., Grimwood J., Davis R.W., Stephens R.S.;  
 RT "Comparative genomes of Chlamydia pneumoniae and C. trachomatis."  
 RL Nat. Genet. 21:385-389(1999).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=AR39;

RA Castagnola P., Gennari M., Morello R., Tonachini L., Marin O.,  
 RA Gaggero A., Cancedda R.;  
 RT "Cartilage associated protein (CASP) is a novel developmentally  
 RT regulated chick embryo protein."  
 RL J. Cell Sci. 110:1351-1359(1997).  
 CC -!- SUBCELLULAR LOCATION: Secreted; extracellular matrix.  
 CC -!- TISSUE SPECIFICITY: Found in articular chondrocytes. Expressed in  
 CC a variety of tissues.  
 CC -!- SIMILARITY: BELONGS TO THE CRTAP / NOS5 FAMILY.  
 CC  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC  
 CC EMBL; X97607; CAA66206.1; -.  
 DR Extracellular matrix; Signal.  
 FT SIGNAL 1 15 POTENTIAL.  
 FT CHAIN 16 271 CARTILAGE ASSOCIATED PROTEIN.  
 FT CARBOHYD 76 76 N-LINKED (GLCNAC...) (POTENTIAL).  
 SQ SEQUENCE 271 AA; 31493 MW; A56343DBF604914C CRC64;  
 Query Match 10.8%; Score 419; DB 1; Length 271;  
 Best Local Similarity 36.1%; Pred. No. 4.2e-20;  
 Matches 100; Conservative 46; Mismatches 107; Indels 24; Gaps 7;  
 Qy 11 TLLAVVAASAAQAEVESEACGNGWTFDILL- ----FAEGTAAAYARDGFWGVLSMERALRSR 65  
 Db 4 TLLAALLATAGAQVERYSFRSPRDELMPLSEAYSRYGLDQYSTENWPESVSYLEVSMRLY 63  
 Qy 66 AALRALRLRCRTCAADFPWELDPDWSPPSQAASGAGALRDLSPFGGLLRRAACLRRC-L 124  
 Db 64 RLLRDEAFCHNCSS- ----AGPLTAPPPAD- ----GELAEALLAGVLRRAQCRLRCKQ 114  
 Qy 125 GPPAAHSUSEMEL- --EPRKRSYNYLOVAFKINKLEKAVAAHTTFVGNPEHMEQON 182  
 Db 115 GLPAPRAQPGRELEEFQRREPKYLOFAYFANNLPKAIKAAHTFLKHPDDEMQRN 174  
 Qy 183 LDYQTMGSKVEADFKDLETQPMQBEFLGLVRLYSEEPQAVPHLEALQOYFVAYEBC 242  
 Db 175 MAYKSIIDAE- HIKDLETKPYENLFVRAVRAINGDNWRTSISDMELALPDPFFKTYDDC 233  
 Qy 243 RALCEGPDYDGVNLYNADLFOAITDHYIQVNLCK 279  
 Db 234 IAACEGSRKIDFK- ----DFYLSIADHYIEVLACK 264



RX MEDLINE=20150255; PubMed=10684935;  
RA Read T.D., Brunham R.C., Shen C., Gill S.R., Heidelberg J.F.,  
RA White O., Hickey E.K., Peterson J., Uterback T., Berry K., Bass S.,  
RA Linher K., Weidman J., Khouri H., Craven B., Bowman C., Dodson R.,  
RA Winn M., Nelson W., DeBoy R., Kolonay J., McClarty G., Salzberg S.L.,  
RA Eisen J., Fraser C.M.,  
RT "Genome sequences of Chlamydia trachomatis MoPn and Chlamydia  
RT pneumoniae AR39.";  
RL Nucleic Acids Res. 28:1397-1406 (2000).  
RN [3]  
RP SEQUENCE FROM N.A.  
RC STRAIN=J138;  
RX MEDLINE=20330349; PubMed=10871362;  
RA Shirai M., Hirakawa H., Kimoto M., Tabuchi M., Kishi F., Ouchi K.,  
RA Shiba T., Ishii K., Hattori M., Kuhara S., Nakazawa T.,  
RT "Comparison of whole genome sequences of Chlamydia pneumoniae J138  
RT from Japan and CWL029 from USA.";  
RL Nucleic Acids Res. 28:2311-2314 (2000).  
RN [4]  
RP SEQUENCE FROM N.A.  
RC STRAIN=TW-183;  
RA Geng M.M., Schumacher A., Muehldorfer I., Bensch K.W., Schaefer K.P.,  
RA Schneider S., Pohl T., Essig A., Marre R., Melchers K.,  
RT "The genome sequence of Chlamydia pneumoniae TW183 and comparison with  
RT other Chlamydia strains based on whole genome sequence analysis.";  
RL Submitted (MAY-2002) to the EMBL/GenBank/DBJ databases.  
CC -!- FUNCTION: The UvrABC repair system catalyzes the recognition and  
CC processing of DNA lesions. UvrC both incises the 5' and 3' sides  
CC of the lesion. The N-terminal half is responsible for the 3'  
CC incision and the C-terminal half is responsible for the 5'  
CC incision (By similarity).  
CC -!- SUBUNIT: Interacts with uvrB in an incision complex (By  
CC similarity).  
CC -!- SUBCELLULAR LOCATION: Cytoplasmic (By similarity).  
CC -!- SIMILARITY: Belongs to the uvrC family.  
CC -!- SIMILARITY: Contains 1 UVR domain.  
CC  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC  
CC -----  
DR EMBL; AE001673; AAD19078.1; -.  
DR EMBL; AE002251; AAF38706.1; -.  
DR EMBL; AP002548; BAA99148.1; -.  
DR EMBL; AE017160; AAP98903.1; -.  
DR PIR; B86608; B86608.  
DR PIR; F72017; F72017.  
DR TIGR; CP0921; -.  
DR HAMAP; MF 00203; -; 1.  
DR InterPro; IPR001943; UvrB/C.  
DR InterPro; IPR004791; UvrC.  
DR InterPro; IPR001162; UvrC.C.  
DR InterPro; IPR00305; UvrC.N.  
DR Pfam; PF01541; Excl\_endo\_N; 1.  
DR ProDom; PD005870; UvrC; 1.  
DR SMART; SM00465; Glyc; 1.  
DR TIGRFAMs; TIGR00194; uvrC; 1.  
DR PROSITE; PS50151; UVR; 1.  
DR PROSITE; PS50164; UVR; 1.  
DR PROSITE; PS50165; UVR-2; 1.  
KW SOS response; Excision nuclease; DNA repair; DNA recombination;  
KW DNA excision; Complete proteome.  
FT DOMAIN 205 240 UVR.  
SQ SEQUENCE 603 AA; 69426 MW; 28017E4FA19FFC84 CRC64;

Query Match 4.18; Score 157.5; DB 1; Length 603;  
Best Local Similarity 20.98; Pred. No. 0.01; 223; Indels 193; Gaps 27;  
Matches 131; Conservative 81; Mismatches 193; Gaps 27;

QY 51 WPGVLSMERALRSRAALRL-----RCRTQCAADFPELDPWSPSPAQASG 100  
DB 111 WPKV-----EAIKTKATSSQRLIFGYPVSAEACHILL-----EVSQWFP----- 152  
QY 101 AGALRDLISFFGGLRRRAAC-----LRRCLGPPAAHSLSEMELEFRKSPNYLQVAFKI 156  
DB 153 ---LRTCSDFEALRKPCILYDMKRLCAPCGVCTPEEYOGTLDK-----AILFLK 201  
QY 157 NKLEKAVAAAHFTFVGNPEHMEMQONLDYYOTMGVKEADFKDLETOPHMQEF----- 209  
DB 202 KIEEVVVDLEKVIQKASDNLFEQAAANYTSLIKQAMAKQOQVEKHFQNIADALGIYR 261  
QY 210 -----RLGVRLYS-ESQPOEAVPHLEAALOEYFA--YECCALCEGYP 250  
DB 262 HKQRTILTLTVRSQKLGARHFFSFFENAQRDQDLSFILQYVVSQPIPKELITPLPL 321  
QY 251 DYDGVNLYENAD---LFQAITDHYIQLV-----NCKQNCVTELASHPSEKPFEDF-- 299  
DB 322 EPTLSIV-LNAESPRLRSPTGIGKELDLATRNKAYAAATL---PSSITLQYQDFON 377  
QY 300 --LPSHYNLOFAYYINIGNYTQAGECAKTYLLFFPFNDEVMMQNLAYAAMLGEBHTRSIG 357  
DB 378 ILRNSQYPIRECYDNA--HMQGAHATGVYIVFENNGFDPKQ-----YRTFSI- 423  
QY 358 PRESAREYRQSRLLLEKELLFPAYDVFGIPFVDPDSWTPPEEVIKELQEKQKSERETAVRI 417  
DB 424 --DSEKTONDLALLEEVL-----RRFHSLLTALPDMIV-----VDGGKTHYNNKTKKI 469  
QY 418 SQEIGNLMKETLVEEKTES--LDVSRLLTREGGPLLYEGISLTMSKLLNGYQRVVMD 475  
DB 470 IQLNLGTGIQVTTAKESNHSRGLNKEKICEFIP---EGFSLPPTSNNLLOFFQ----- 521  
QY 476 GVISDHECOELQRLTNVAATSGDGYRGQTSPTNEKPYGVTVFKALKLQGEQKVPLOSA 535  
DB 522 -ILRD-----EAHFAISKH---RKRGKALFEQEKIPGIGEV----- 555  
QY 536 HLYNVNTEKVRIMESYFRLDTPLVFSYSHLVCTAIEBQVQAKRDSHPVVDNCILNA 595  
DB 556 -----KRRLLQK-----FKSWKQVMSQBELEA----- 580  
QY 596 ETLVCVKEPAYTRDYSAILYNGDFD 623  
DB 581 -----IPGLTKKDIAVLLARQKDFN 600  
  
RESULT 7  
EG27\_CABEL  
ID EG27\_CABEL STANDARD; PRT; 1129 AA.  
AC Q09228; Q09229; Q8MOP3; Q9XYD0;  
DT 01-OCT-1996 (Rel. 34, Created)  
DT 28-FEB-2003 (Rel. 41, Last sequence update)  
DT 28-FEB-2003 (Rel. 41, Last annotation update)  
DE Egg laying defective protein 27.  
GN EGL-27 OR C04A2.3.  
OS Caenorhabditis elegans.  
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;  
OC Rhabditidae; Peloderinae; Caenorhabditis.  
OX NCBI\_TaxID=6239;  
RN [1]  
RP SEQUENCE FROM N.A. (ISOFORM A), FUNCTION, SUBCELLULAR LOCATION, AND  
RP DEVELOPMENTAL STAGE.  
RC STRAIN=Bristol N2;  
RX MEDLINE=99128194; PubMed=9927605;  
RA Herman M.A., Ch'ng Q., Hattenbach S.M., Ratliff T.M., Kenyon C.,  
RA Herman R.K.,  
RT "EGL-27 is similar to a metastasis-associated factor and controls cell  
RT polarity and cell migration in C. elegans.";  
RL Development 126:1055-1064 (1999).  
RN [2]  
RP SEQUENCE FROM N.A.  
RC STRAIN=Bristol N2;  
RA Du Z.,  
RL Submitted (JUL-1995) to the EMBL/GenBank/DBJ databases.



RX MEDLINE=92335263; PubMed=16311106;  
RA Byers T.J., Brandin E., Lue R., Winograd E., Branton D.;  
RT "The complete sequence of Drosophila beta-spectrin reveals  
RT supra-motifs comprising eight 106-residue segments.";  
RL Proc. Natl. Acad. Sci. U.S.A. 89:6187-6191(1992).  
RN [2]  
RP SEQUENCE OF 1-800 FROM N.A.  
RX MEDLINE=90009037; PubMed=2677025;  
RA Goldstein L.S.B.;  
RA "Sequence similarity of the amino-terminal domain of Drosophila beta  
RT spectrin to alpha actinin and dystrophin.";  
RL J. Cell Biol. 109:1633-1641(1989).  
RN [3]  
RP SEQUENCE FROM N.A.  
RX STRAIN=Berkley;  
RX MEDLINE=20196006; PubMed=107311132;  
RA Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,  
RA Ananides P.G., Scherer S.E., Li P.W., Hostkins R.A., Galie R.F.,  
RA George R.A., Lewis S.B., Richards S., Ashburner M., Henderson S.N.,  
RA Sutton G.G., Wortman J.R., Vandeil M.D., Zhang Q., Chen L.X.,  
RA Branton R.C., Rogers Y.-H.C., Blazej R.G., Champe M., Pfeiffer B.D.,  
RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,  
RA Abril J.F., Agbayani A., An H.-J., Andrews-Pfannkuch C., Baldwin D.,  
RA Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,  
RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,  
RA Borkova D., Botchan M.R., Bouck J., Brokstein P., Brotter P.,  
RA Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,  
RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,  
RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,  
RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,  
RA Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,  
RA Flosler C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,  
RA Glodex A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,  
RA Harris N.L., Harvey D.A., Heiman T.J., Hernandez J.R., Houck J.,  
RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,  
RA Jalali M., Kalush P., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,  
RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,  
RA Lasko P., Lei Y., Levitsky A.A., Li J.H., Li Z., Liang Y., Lin X.,  
RA Liu X., Mattei Y., McIntosh T.C., McLeod M.P., McPherson D.,  
RA Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,  
RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,  
RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacle J.M.,  
RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,  
RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,  
RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,  
RA Spier E., Spradling A.C., Stapieton M., Strong R., Sun E.,  
RA Svirskaas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,  
RA Wang Z.-Y., Wassarman D.A., Weinstein G.M., Weissenbach J.,  
RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,  
RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,  
RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Smith H.O.,  
RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;  
RT "The genome sequence of Drosophila melanogaster.";  
RL Science 287:2185-2195(2000).  
RN [4]  
RP CHARACTERIZATION.  
RX MEDLINE=88059242; PubMed=3680372;  
RA Dubreuil R., Byers T.J., Branton D., Goldstein L.S.B., Kiehart D.P.;  
RT "Drosophila spectrin. I. Characterization of the purified protein.";  
RL J. Cell Biol. 105:2093-2102(1987).  
RN [5]  
RP STRUCTURE BY NMR OF 2145-2262.  
RX MEDLINE=96164435; PubMed=8591029;  
RA Zhang P., Talluri S., Deng H., Branton D., Wagner G.;  
RT "Solution structure of the pleckstrin homology domain of Drosophila  
RT beta-spectrin.";  
RL Structure 3:1185-1195(1995).  
CC -1- FUNCTION: Spectrin is the major constituent of the cytoskeletal  
CC network underlying the erythrocyte plasma membrane. It associates  
CC with band 4.1 and actin to form the cytoskeletal superstructure of  
CC the erythrocyte plasma membrane. Interacts with calmodulin in a  
CC calcium-dependent manner.

CC -1- SUBUNIT: Native spectrin molecule is a tetramer composed of two  
CC antiparallel heterodimers joined head to head so that each end  
CC of the native molecule includes the C-terminus of the alpha  
CC subunit and the N-terminus of the beta subunit.  
CC -1- SIMILARITY: Belongs to the spectrin family.  
CC -1- SIMILARITY: Contains 2 calponin-homology (CH) domains.  
CC -1- SIMILARITY: Contains 1 PH domain.  
CC -1- SIMILARITY: Contains 17 spectrin repeats.  
CC  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC  
CC EMBL; M92288; AAA28399.1; --  
CC EMBL; AE003506; AAF48751.1; --  
CC PIR; A46147; A46147.  
CC PDB; 1DRO; 03-APR-96.  
CC FlyBase; FBgn0003471; beta-Spec.  
CC GO; GO:0045169; C:fusome; IDA.  
CC GO; GO:0005886; C:plasma membrane; IDA.  
CC GO; GO:0045170; C:spectrosome; IDA.  
CC GO; GO:0008017; F:microtubule binding; IDA.  
CC GO; GO:0007274; P:neuromuscular synaptic transmission; IMP.  
CC InterPro; IP0001589; Actbind actinin.  
CC InterPro; IP0001715; Calponin-like.  
CC InterPro; IP0001849; PH.  
CC InterPro; IP0002017; Spectrin.  
CC InterPro; IP0001605; Spectrin\_PH.  
CC Pfam; PF00307; CH; 2.  
CC Pfam; PF00169; PH; 1.  
CC Pfam; PF00435; spectrin; 17.  
CC PRINTS; PR00683; SPECTRINPH.  
CC SMART; SM00033; CH; 2.  
CC SMART; SM00233; PH; 1.  
CC SMART; SM00150; SPEC; 17.  
CC PROSITE; PS00019; ACTININ\_1; 1.  
CC PROSITE; PS00020; ACTININ\_2; 1.  
CC PROSITE; PS00021; CH; 2.  
CC PROSITE; PS00003; PH DOMAIN; 1.  
CC Cytoskeleton; Membrane; Calmodulin-binding; Actin-binding; Repeat;  
CC 3D-structure; Actin capping.  
FT DOMAIN 1 271 ACTIN-BINDING (BY SIMILARITY).  
FT DOMAIN 50 154 CH 1.  
FT DOMAIN 169 271 CH 2.  
FT REPEAT 298 408 SPECTRIN 1.  
FT REPEAT 418 522 SPECTRIN 2.  
FT REPEAT 524 633 SPECTRIN 3.  
FT REPEAT 635 739 SPECTRIN 4.  
FT REPEAT 741 844 SPECTRIN 5.  
FT REPEAT 846 950 SPECTRIN 6.  
FT REPEAT 952 1057 SPECTRIN 7.  
FT REPEAT 1059 1167 SPECTRIN 8.  
FT REPEAT 1169 1273 SPECTRIN 9.  
FT REPEAT 1275 1378 SPECTRIN 10.  
FT REPEAT 1380 1485 SPECTRIN 11.  
FT REPEAT 1487 1591 SPECTRIN 12.  
FT REPEAT 1593 1697 SPECTRIN 13.  
FT REPEAT 1699 1804 SPECTRIN 14.  
FT REPEAT 1806 1910 SPECTRIN 15.  
FT REPEAT 1912 2016 SPECTRIN 16.  
FT REPEAT 2018 2078 SPECTRIN 17.  
FT DOMAIN 2147 2259 PH.  
FT CONFLICT 2278 2278 D -> Y (IN REF. 2).  
FT STRAND 2150 2157  
FT TURN 2166 2167  
FT STRAND 2173 2179  
FT STRAND 2185 2187  
FT HELIX 2192 2195  
FT TURN 2197 2198

```

FT STRAND 2206
FT TURN 2215
FT TURN 2216
FT TURN 2224
FT STRAND 2226
FT STRAND 2230
FT STRAND 2236
FT STRAND 2240
FT HELIX 2245
FT TURN 2258
FT TURN 2259
SQ SEQUENCE 2291 AA; 265737 MW; 5CDFB0C548BBC39B CRC64;

Query Match
  3.1%; Score 120.5; DB 1; Length 2291;
Best Local Similarity 19.0%; Pred. No. 16;
Matches 121; Conservative 96; Mismatches 236; Indels 183; Gaps 27;

Qy 175 EHMEMQNLDYQTMGSGVGEADPKDLETQPHMOEFLGVLRYSE---EQPQ-----EAV 225
Db 1100 QHQSIREIDNY-----TEDYK-----NMVEY---GERLTSEGSTSDDPQYMFRLERL 1144
Qy 226 PHLEAALQYFVAVEECRACALCEGPDYDGYNYLEYNADLFQAITDHYIQLVNCQKCVTE 285
Db 1145 NALKDGEELHOMWENQVLLSQSLDQQLFNDRARQTEVLLSQEHLFS-----KDDTPVN 1200
Qy 286 LASHPSREKPFEDLPSPHYNLOPAYYNYGTQAGECAKTYLLFPNDWMMQNLAYYA 345
Db 1201 LEQAENQKHEAFLTT-----MEANDDKINTLLQVAD 1233
Qy 346 ANLGEH--TRSGPRES-----AKEVRSRLLEKELL-----PFAYDV-----382
Db 1234 TLVEKDFADKTKGRAENTIGTRDDNRQALDQHEKLNQVXKLHFEFLQDLEELAEWVOE 1293
Qy 383 -----FGIPFVDPDSTPEVI-----PKLQEKQKSERETAVRISQEIENLAK 426
Db 1294 KYATSQDESYSRATKTHQAPAEATAANKERLFEAKSAQ-----LSKEKEPKD 1349
Qy 427 EITLVEKYESLDVSRRLTREGGLPYEGISITMSKLLNGYQVVMQGVISDHECQEL 486
Db 1350 VIEPKLAKAQFDLEVHTKEKAMLFDA-----NRELVQOTCDIDSIYITDEKQ-- 1402
Qy 487 QRLTNVAATSGDVGRTGSPHTNEKPYGVTVFKALKLGQEGKVPQSAHLYYNVTE--- 543
Db 1403 -----IVSGDANDLTSNLMQKQVITQTMVAKARQVEIDKQTEYLQXTVPEEKI 1455
Qy 544 -----KVRIMESYRLDTPLYSYSLVCTRTAIEVQA-ERKDDSHPVVDNCILNAET 597
Db 1456 EPIVVKTAVERPEKIKAPL-----LERQKALEKKEAFQCRD---VEDEK 1500
Qy 598 LVCVKEPPATFRDYSAILYNGDFDGGNFYFTELDKATVTAEVQPCQGRVGFSSGTEN 657
Db 1501 LWDEKLPVANGPDYGNLSF-----NVHVLKKNQSLATEID-----NHE 1540
Qy 658 PHGVKAVTRQRCALALWFTLDPRHSDRVQA--DDLV-----KMLFSPE 701
Db 1541 PR-INAICNNRKLJ-----DEGHEDAKKFEALISDLTKQWELKDAIENRRKHLSES 1593
Qy 702 EMDLSQEPIDAOQGPPEPAQESI--SGSESQPKDEL 736
Db 1594 KV---QQYFFDAQEASWSEQLYMMVDRGXDEI 1626

RESULT 9
CEP2_MOUSE
ID CEP2_MOUSE STANDARD; PRT; 609 AA.
AC Q60952;
DT 10-OCT-2003 (Rel. 42, Created)
DT 10-OCT-2003 (Rel. 42, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Centrosomal protein 2 (Intranuclear matrix protein) (Fragment).
GN CEP2 OR INMP.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OC NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.

```

```

RC TISSUE=Macrophage;
RX MEDLINE=97107429; PubMed=8950169;
RA Menz K., Radomski N., Jost E.;
RT "INMP, a novel intranuclear matrix protein related to the family of
RT intermediate filament-like proteins: molecular cloning and sequence
RT analysis.";
RL Biochim. Biophys. Acta 1309:14-20(1996).
CC -!- FUNCTION: Probably plays an important role in centrosome cohesion
CC during interphase (By similarity).
CC -!- SUBUNIT: Monomer and homodimer (Probable). Forms a complex in
CC vitro with both NEK2 kinase and the PP1CC catalytic subunit of
CC protein phosphatase 1 (PP1) (By similarity).
CC -!- SUBCELLULAR LOCATION: Present in perinuclear region. Component of
CC the core centrosome (By similarity). Some presence of the protein
CC in the nucleus is not excluded.
CC -!- PTM: C terminal part is phosphorylated by NEK2. It is
CC dephosphorylated in vitro by the PP1 phosphatase (By similarity).
CC -!- CAUTION: Ref.1 sequence differs from that shown due to frameshifts
CC in positions 37, 108, 118, 536, 565 and 567.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; U33198; AAB41824.1; ALT_FRAME.
KW MGD; MGI:108084; Cep2.
KW Coiled coil; Phosphorylation.
FT NON_TER 1 1
FT DOMAIN 1 395 COILED COIL (POTENTIAL).
FT DOMAIN 426 487 COILED COIL (POTENTIAL).
FT DOMAIN 513 546 COILED COIL (POTENTIAL).
FT DOMAIN 43 289 GLU-RICH.
SQ SEQUENCE 609 AA; 69170 MW; 13CEC554AE018103 CRC64;

Query Match
  3.1%; Score 118.5; DB 1; Length 609;
Best Local Similarity 22.5%; Pred. No. 3.4; Indels 121; Gaps 25;
Matches 111; Conservative 70; Mismatches 191;

Qy 132 LSEMELEFRK-RSPYNYLQVAYFKINKLEKAVAAHTFFVGNPEHMEMQNLDYQTM 190
Db 58 LREELEVEGRQVRALEEVLGDLRAESREHEKAVLA-----LQORCAEQAEH 104
Qy 191 GVKEADFKD--LETQPHM--QEFRLGV-----RLYSEEP-----OEAVPHLEAL 232
Db 105 EAEARTQDSWLQARPTLITEQEQLALRAENQYRRQEEAANVSQAALQALSKAQAAL 164
Qy 233 QBYFVAYECCALCEGPDYDGYNYLEYNADLFQAITDHYIQLVNCQKCVTELASHPSR 292
Db 165 QEKQSLLEQAELS-----HTLEASTAALQATLD-----TCQ-----AR 198
Qy 293 EKPFEDFLPSHYNLO---PAYNYNYGTQAGECAKTYLLFFPNDEVMNQNLAYYAALMG 349
Db 199 ARQLEEARLRREGIQAALQHQHEVTOHQELCOK-----EELRLLEKAGARRS 250
Qy 350 EEHTRSIGPRESAXEYRQ---RSLLE--KEL-LPFAYDVFQIPFVDPDSTPEVIP--- 400
Db 251 QEN--GIQEKQSLQEQEETRRLLLESIKELQLTVAQEEELMLREASSPSHRALEAK 308
Qy 401 ---KRLQEKQKSER-ETAVRISQEIGNLMKEIEFTLVEEKESLDVSRRLTREGGLPYEG 456
Db 309 PALQFLPAQQLERLQALROTE-----AREIEW--REKAQDL-----A 345
Qy 457 ISLTWNSKLLNGYQVVM--DGVISDHCEQBLQRLTNVAATSGDVGRTGTS--PHTNEK 512
Db 346 LSLAQSKASISLQEIITWFLQASVLERE-SEQRLQELVLSQALBEQSGGHSISRA 404
Qy 513 FYGVTVFKALKGQEGKVPQSAHLYYNVTEKVRIMESYRLDTPLYFSYSHLVC----- 569
Db 405 DQGPVKQESQFGEVETEPSPGVEKERLTQRLERLQQAVALLEV-----DRSKLQCHNAQ 460

```



FT DOMAIN 1011 1131 PH.  
FT DOMAIN 1160 1457 SH3-BINDING (POTENTIAL).  
FT SITE 1521 1526 SH3-BINDING (POTENTIAL).  
SQ SEQUENCE 1597 AA; 18344 MW; 7B7286C2305676DA CRC64;

Query Match 3.0%; Score 117; DB 1; Length 1597;  
Best Local Similarity 20.5%; Pred. No. 14;  
Matches 120; Conservative 88; Mismatches 190; Indels 186; Gaps 28;

QY 132 LSEMELEFRKSPYNYLVQVAYFKINKLEKAVAAHTFFVGN---PEHM--EMOQNLDYY 186  
DB 343 LSEQKAMINAMDSKIRSLERIVSELSANKLANSSILFTQNNKKAQEMISELRQKPYL 402  
QY 187 QTMGKVEADFKDLET-----QPHMOEFRL-----GVRLYSEEPQOE---AVPHL 228  
DB 403 ETQAGKLEAQNVRKLEQLEKISHQDHSKRLLETRLEVSLEHEEQKLELRQKTEL 462  
QY 229 EALQOE---YFVAYEECRALCEGPDYDGYNYLEYNADL---FOAITDHYIOLN---C 278  
DB 463 QLSQERESQITALQAPRAALESQLRQAKTELETTABEEIOALTARDEIQRKFDAL 522  
QY 279 KQNC--VTLE-----ASHPSREKPFEDF-----LPSHNYLYOFAY- 311  
DB 523 RNSCTVITDLEEQNLQNTEDNAELNNQNFYLSKQLOEASGANDEIVQLRSEVDHLRREIT 582  
QY 312 ---YNIQNYTQAGECAKTYLLFFP-----NDEVNQNLAYA--AMLGEHT--- 353  
DB 583 EREMQLTSQKOTMEALKTTCTMLEEQVLDLEALNDELLEKERQWEAWRSVLGDEKSOPEC 642  
QY 354 ---RSIGPRE-----SAKEYR-----QRSLEKEKELLFPAYDV 382  
DB 643 RVRELQRLMDEKQSRARADQRTESQVVELAVKHKABAILAQALKEQKL----- 695  
QY 383 FGIPFVDPDSWTP-----EVIKRLQKQKSERETAVRSQEIQNLKMKIET- 430  
DB 696 -----KAESLDKLNLEKHAMLENARSLOQKLETERELKQRLLEEQAKLQOQMDLQ 749  
QY 431 ---LVEEKTESLDVSRLTR-EGGPLLYEGISLTWNSKLLGYORVVMGVDGVISDHECQ 484  
DB 750 KWHIFRITQGLQALDADLADLLKTERSDEYQ-----LENIQVLSHEKVMGFTSQ--- 801  
QY 485 ELQRLTNVAATSGDYGRTSPHTPNKFGYVTVFKALKLQO---EGKVPLOSAHYLYNV 541  
DB 802 -----QTK-----LIDFLQAKMDQPAKKKVPLOYNELKAL 833  
QY 542 TEKVRIMESYRLDTPLYSYSHLVCRATAIEEVAERKDDSH 585  
DB 834 EKEKARCAE-----LEAL--OKTRIELRSAREEA-AHRKATDHP 870

RESULT 12

CTRO HUMAN STANDARD; PRT; 1286 AA.  
AC Q4578; QUP27;  
DT 30-MAY-2000 (Rel. 39, Created)  
DT 30-MAY-2000 (Rel. 39, Last sequence update)  
DT 28-FEB-2003 (Rel. 41, Last annotation update)  
DE Citron protein (Rho-interacting, serine/threonine kinase 21)  
DE (fragment).  
GN CIT OR STK21 OR KIRA0949.  
OS Homo sapiens (Human).  
OC Sukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
OX NCBI\_Taxid:9606;  
[1]  
RN SEQUENCE FROM N.A.  
RP Connell M., Goela D., Harper M.;  
RL Submitted (OCT-1997) to the EMBL/GenBank/DBJ databases.  
[2]  
RC  
RP  
RC  
RX TISSUE=Brain;  
RX MEDLINE=9246063; PubMed=10231032;  
RA Nagase T., Ishikawa K.-I., Suyama M., Kikuno R., Hirose M.,

Query Match 3.0%; Score 117; DB 1; Length 1286;  
Best Local Similarity 20.5%; Pred. No. 12;  
Matches 120; Conservative 88; Mismatches 190; Indels 186; Gaps 28;

QY 132 LSEMELEFRKSPYNYLVQVAYFKINKLEKAVAAHTFFVGN---PEHM--EMOQNLDYY 186  
DB 34 LSEQKAMINAMDSKIRSLERIVSELSANKLANSSILFTQNNKKAQEMISELRQKPYL 93  
QY 187 QTMGKVEADFKDLET-----QPHMOEFRL-----GVRLYSEEPQOE---AVPHL 228  
DB 94 ETQAGKLEAQNVRKLEQLEKISHQDHSKRLLETRLEVSLEHEEQKLELRQKTEL 153  
QY 229 EALQOE---YFVAYEECRALCEGPDYDGYNYLEYNADL---FOAITDHYIOLN---C 278  
DB 154 QLSQERESQITALQAPRAALESQLRQAKTELETTABEEIOALTARDEIQRKFDAL 213  
QY 279 KQNC--VTLE-----ASHPSREKPFEDF-----LPSHNYLYOFAY- 311  
DB 214 RNSCTVITDLEEQNLQNTEDNAELNNQNFYLSKQLOEASGANDEIVQLRSEVDHLRREIT 273  
QY 312 ---YNIQNYTQAGECAKTYLLFFP-----NDEVNQNLAYA--AMLGEHT--- 353  
DB 274 EREMQLTSQKOTMEALKTTCTMLEEQVLDLEALNDELLEKERQWEAWRSVLGDEKSOPEC 333

```
QY 354 -----RSIGPRE-----SAKEYR-----QRSLEKELLFPAYDV 382
Db 334 RVRELQMLDTEKOSRARADQITEGROVVVELAVKHEKAEILALOQALKEQKL----- 386
QY 383 FGIPFVDPDSWTPE-----EVIKRLQEKOKSERETAVRISQIGNLMKSIET- 430
Db 387 -----KAESLSDKLNDLEKKHAMLEVNARSLOKLETERELKQRLLEQAKLQOQMDLQ 440
QY 431 -----LVEETKESLDVSRUTR-EGGPLYEGISLWNKSLNGYQVVDGVISDHCO 484
Db 441 KNIHFRITQLOALDRADLLKTERSDEYQ-----LENIQVLYSHEKVMEGTISQ----- 492
QY 485 ELQRLTNVAATSGDVGRTSPHTPNEKFGVTVFVKALQGO-----RGKVPLOSAHLYNVV 541
Db 493 -----QTK-----LDFLOAKMDQPAKKKVPLOQNELKAL 524
QY 542 TEKVRIIMESYFLDTPLYFSYSHLVCRTAIEVQAEKDDSH 585
Db 525 EKEKARCAE-----LEAL--QKTRIELRSAREEA-AHRKATDHP 561

RESULT 13
RADO_SULSO
ID RA50_SULSO STANDARD; PRT; 864 AA.
AC Q97WH0;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE DNA double-strand break repair rad50 ATPase.
GN RAD50 OR S502249.
OS Sulfolobus solfataricus.
OC Archaea; Crenarchaeota; Thermoprotei; Sulfolobales; Sulfolobaceae;
OC Sulfolobus.
OX NCBI_TaxID=2287;
RN [1]
SEQUENCE FROM N.A.
RC STRAIN=ATCC 35052 / DSM 1617 / P2;
RX MEDLINE=21332296; PubMed=11427726;
RA She Q., Singh R.K., Confalonieri F., Zivanovic Y., Allard G.,
RA Awavez M.J., Chan-Weher C.C.-Y., Clausen I.G., Curtis B.A.,
RA De Moors A., Ernauso G., Fletcher C., Gordon P.M.K.,
RA Heikamp-de Jong I., Jeffries A.C., Kozera C.J., Medina N., Peng X.,
RA Thi-Ngoc H.P., Redder P., Schenk M.E., Theriault C., Tolstrup N.,
RA Charlebois R.L., Doolittle W.F., Duguet M., Gaasterland T.,
RA Garrett R.A., Ragan M.A., Senses C.W., Van der Oost J.;
RT "The complete genome of the crenarchaeon Sulfolobus solfataricus P2.";
RL Proc. Natl. Acad. Sci. U.S.A. 98:7835-7840(2001).
CC -!- FUNCTION: Involved in DNA double-strand break repair (DSBR). The
CC rad50/mrell complex possesses single-strand endonuclease activity
CC and ATP-dependent double-strand-specific exonuclease activity.
CC Rad50 provides an ATP-dependent control of mreII by unwinding
CC and/or repositioning DNA ends into the mreII active site (By
CC similarity).
CC -!- SUBUNIT: Forms a complex with mreII (By similarity).
CC -!- SIMILARITY: Belongs to the SMC family. RAD50 subfamily.
CC
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; AEO06829; AAK42417.1; -.
CC PIR; B90395; B90395.
CC HAMAP; MF 00449; -.
CC InterPro; IPR003439; ABC transporter.
CC InterPro; IPR007517; Rad50_zn_hook.
CC InterPro; IPR003405; SMC_C.
CC InterPro; IPR003395; SMC_N.
CC Pfam; PF04423; Rad50_zn_hook; 1.
CC Pfam; PF02483; SMC_C; 1.
```

```
DR Pfam; PF02463; SMC_N; 1.
DR ProDom; PD00006; ABC_transporter; 1.
KW DNA repair; Hydrolase; ATP-binding; Coiled coil; Complete proteome.
FT NP_BIND 30 37 ATP (BY SIMILARITY).
ET DOMAIN 164 701 COILED COIL (POTENTIAL).
SQ SEQUENCE 864 AA; 101601 MW; 657076AA9B709FC CRC64;

Query Match
Best Local Similarity 18.3%; Pred. NO. 8.1;
Matches 128; Conservative 103; Mismatches 227; Indels 242; Gaps 29;

QY 141 RKSPY-----NYLOVAVFKINKLEKAVAAAHATFFVGNPEHMEMQONLDYYQTMGSVKREAD 196
Db 232 KRGNYLKLTTTLKIKEGELNELNRSI-----EELRKQTEMDQLE----- 272
QY 197 FKDLTQPMQEFRLGVRLY-----SEEQPOEAVPHLEAALQYFVAYEECRALCEGYD 251
Db 273 -KEINELNLRNLIKLPKPEYVLAKSHTEMSANVINLEKEIEEYKAIKRRKEEL---EPK 328
QY 252 YDGYNYLEYNADLFOAITDHYIQV-----LNKQNC-----VTELASH-----PSREKFFE 297
Db 329 YLYKELERKLEELQPKYQYLYKLKSLDLSKLNLERLEKDASELSNDIDKVNLSLEQKVE 388
QY 298 DFLPSHVNLYQFAYVYNYGNYTQAGECAKTYLLFFPNDVNMN-----ONLAYYAALMGE 350
Db 389 ETRKKQLNL-----RAQLAKVESLISEKNEIINNISOVEGETCPVGRPLDE 435
QY 351 EHTRSIGPRESAXYRQSRSLLEKELLFFAYDVFGIPFVDPDSWTPPEVPIKRLQEKOKSE 410
Db 436 EHKQKI--IKEAKSYILQLELNKNEI-----EBELKKITNELNKIE 474
QY 411 RE-----TAVRISOEIGNLMKEITL--VVEK----- 435
Db 475 REYRLSNKASYDNVWRQLKNEELENLSELESKNDEEEKKINEEVKELKLYEE 534
QY 436 -----TKESLDVSRLTREGGLLYEGISLTM-----NSKLLN-GYQRV 472
Db 535 FWRLSKYTKELDKKRVKLDKEMKKKKEIEKXRGLESELKGLDRKALESKILDENKRV 594
QY 473 VMD-----GVISH-----ECQELQRLTNVAATSGDGYRGQTSPTPNKFGVGT 517
Db 595 KLDKEMKKKGLIEDYIRQVKLLQEEVNNLREEVNIIQFDENRY----- 637
QY 518 VFKALKLQEGKVPLOSAHLYNVYVTEKVRRIIMESYFLDTPLYFSYSHLVCRTAIEEV-- 575
Db 638 -----NELKTSLDA-----YNLSLKEKENKS--RIEGELE-----SLEKDIIEISN 677
QY 576 -----QAEKDDSHPVHVDNCLNAETLVCVKEPPATVTRDYSAILYN-----GDFDGG 625
Db 678 RIANYELQLKDREKIIINAKNLEKIRGALGERKLQSYIIMTTKOLIENLNLDIISKFD-- 735
QY 626 NFYFTELDAKTVTAEVQPOCGRA-----VGFSSGTENPHGVKAVTRGQRCALAL-- 674
Db 736 -----LSIKKVEIMEPKTGRGRSSSGDILVYNSGDTLP--IVSLSGGERIALSLALR 787
QY 675 -----WFTLDPHRSERDRVQADDLVKMLFSPPEE 702
Db 788 LAIAKALMSNTNFFILDEPTIHLDDQKAYLIEIIRAKE 827

RESULT 14
ADDB_BACSU
ID ADDB_BACSU STANDARD; PRT; 1166 AA.
AC P23477;
DT 01-NOV-1991 (Rel. 20, Created)
DT 01-NOV-1991 (Rel. 20, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE ATP-dependent nuclease subunit B.
GN ADDB OR B5U10620.
OS Bacillus subtilis.
OC Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.
OX NCBI_TaxID=1423;
RN [1]
```



RP SEQUENCE FROM N.A.  
RC STRAIN=OG1;  
RX MEDLINE=91267926; PubMed=1646786;  
RA Kocisträ J., Venema G.;  
RT "Cloning, sequencing, and expression of *Bacillus subtilis* genes  
RT involved in ATP-dependent nuclease synthesis.";  
RL J. Bacteriol. 173:3644-3655(1991).  
[2]  
RN SEQUENCE FROM N.A.  
RP STRAIN=168;  
RX MEDLINE=98240224; PubMed=9579061;  
RA Noback M.A., Holsappel S., Kiewiet R., Terpstra P., Wambutt R.,  
RA Wedler H., Venema G., Bron S.;  
RT "The 172 kb *prkA*-*addAB* region from 83 degrees to 97 degrees of the  
RT *Bacillus subtilis* chromosome contains several dysfunctional genes,  
RT the *glyB* marker, many genes encoding transporter proteins, and the  
RT ubiquitous hit gene.";  
RL Microbiology 144:859-875(1998).  
[3]  
RN SEQUENCE FROM N.A.  
RP STRAIN=168;  
RX MEDLINE=98044033; PubMed=9384377;  
RA Kunst F., Ogasawara N., Moszer I., Albertini A.M., Alloni G.,  
RA Azevedo V., Bertero M.G., Bessieres P., Bolotin A., Borchert S.,  
RA Borries R., Boursier L., Brans A., Braun M., Brignell S.C., Bron S.,  
RA Brouillet S., Bruschi C.V., Caldwell B., Capuano V., Carter N.M.,  
RA Choi S.K., Codani J.J., Connerton I.F., Cummings N.J., Daniel R.A.,  
RA Denizot F., Devine K.M., Dusterhoft A., Ehrlich S.D., Emmerson P.T.,  
RA Ertan K.D., Errington J., Fábret C., Ferrari E., Foulger D.,  
RA Frim C., Fujita M., Fujita Y., Fuma S., Galizzi A., Galleron N.,  
RA Gitz S., Glaser P., Coffeau A., Golightly E.J., Grandi G.,  
RA Giuseppe G., Guy B.J., Haga K., Halech J., Harwood C.R., Henaut A.,  
RA Hilbert H., Holsappel S., Hosono S., Hullo M.P., Itaya M., Jones L.,  
RA Joris B., Karamata D., Kasahara Y., Klaerr-Blanchard M., Klein C.,  
RA Kobayashi Y., Koetter P., Koningsstein G., Krogh S., Kumano M.,  
RA Kurita K., Lapinus A., Lardinois S., Lauber J., Lazarevic V.,  
RA Lee S.M., Levine A., Liu H., Masuda S., Mauel C., Medigue C.,  
RA Medina N., Mellado R.P., Mizuno M., Moestl D., Nakai S., Noback M.,  
RA Noone D., O'Reilly M., Ogawa K., Ogiwara A., Oudega B., Park S.H.,  
RA Parro V., Pohl T.M., Portelle D., Porwollik S., Prescott A.M.,  
RA Prescan E., Pujic P., Purnelle B., Rapoport G., Rey M., Reynolds S.,  
RA Rieger M., Rivolta C., Rocha E., Roche B., Rose M., Sadie Y.,  
RA Sato T., Scanlan E., Schleich S., Schroeter R., Scoffone F.,  
RA Sekiguchi J., Sekowska A., Seror S.J., Serror P., Shin B.S., Soldo B.,  
RA Sorokin A., Tacconi E., Takagi T., Takahashi H., Takemaru K.,  
RA Takeuchi M., Tamakoshi A., Tanaka T., Terpstra P., Tognoni A.,  
RA Tosato V., Uchiyama S., Vandenbol M., Vannier F., Vassarotti A.,  
RA Viari A., Wambutt R., Wedler H., Wedler H., Weitzengger T.,  
RA Winters P., Wipat A., Yamamoto H., Yamane K., Yasumoto K., Yata K.,  
RA Yoshida K., Yoshikawa H.F., Zumbstein E., Yoshikawa H., Danchin A.;  
RT "The complete genome sequence of the Gram-positive bacterium *Bacillus subtilis*.";  
RL Nature 390:249-256(1997).  
CC -!- FUNCTION: THE ENZYME COMPLEX MAY HAVE A WIDE VARIETY OF CATALYTIC  
CC ACTIVITIES INCLUDING ATP-DEPENDENT EXONUCLEASE, ATP-STIMULATED  
CC ENDONUCLEASE, ATP-DEPENDENT HELICASE AND DNA-DEPENDENT ATPASE  
CC ACTIVITIES.  
CC -!- SUBUNIT: THE B SUBUNIT IS ATP-DEPENDENT NUCLEASE COMPLEX IS FORMED  
CC BY ONE B SUBUNIT IN CONJUNCTION WITH ONE A SUBUNIT.  
CC  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed, usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC  
CC EMBL; M63489; AAA22200.1; -;  
CC EMBL; Y14081; CAA74481.1; -;  
CC EMBL; Z99109; CAB12902.1; -;  
CC PIR; A39432; A39432.  
CC Subtilist; BG10465; addB.

DR InterPro; IPR000212; UvrD-helicase.  
DR Pfam; PF00580; UvrD-helicase; 1.  
KW Hydrolase; Nuclease; Exonuclease; Endonuclease; Helicase; ATP-binding;  
KW DNA repair; Complete proteome.  
FT NP\_BIND 8 15 ATP (POTENTIAL).  
SQ SEQUENCE 1166 AA; 134631 MW; 1A57BBE81A08AB4A CRC64;  
Query Match 3.0%; Score 116; DB 1; Length 1166;  
Best Local Similarity 19.1%; Pred. No. 12;  
Matches 134; Conservative 95; Mismatches 230; Indels 244; Gaps 33;  
QY 127 PAASLSSEMELEERKSPYNYLVQAVFKINKLEKAVAAAHATFFVGNPEHMEQNLIDY- 185  
DB PSYREPEHELEL-FRMTCGK-----TYRLHQKAK-----ELNLDIT 274  
QY 186 YQTMG-----VKEADFKDLETPHMQBFLGLVRLYSEEQCEAVPHLEAA- 231  
DB YKELSGTERHTKTPELAHLAQ---YEARPAIP-YAEK--QEALTVMQANRRAELEGIA 328  
QY 232 -----LQYFYVAYEECPALCEGYDYDGY---NYLEYNADLFOAITDHYIQLVCKQNC 282  
DB REIHALVREKGYRYKDVAILARQEPEDYKMDYKVFADYEIPIYF---IDGKASMLN----- 380  
QY 283 VTELASHPESREKPFEDFLPSHYNYLVQFAYYVGN--YQAGECAKTVLLFPFN----- 333  
DB -----HPLIEFIRSSLDVLK-----GNWRYEAVFRCVKTLLPLNEPKAKVR 423  
QY 334 ---DEVNQNLAAYVAMLGEEHRSIGRESAKY----- 365  
DB 424 EOVDOLENYCIAY--GIKGRWTK--GDRFOYRFFVSLDDDDFAQTQDQIEIEMNNMLNTRD 479  
QY 366 -----RQSLLEKELLFFAYDVFGIPFVDPDSWTSWTEEVIPKRL-CEKOKSE 410  
DB 480 WIVPLFOLQKMKKAKIVQKAEALRY-----LEETDVLKLDQERQAE 526  
QY 411 RETAVRISQEIENLMKEIETIWE-----KTKESLDSVRLTREGGPLLVEGISLTWNSK 464  
DB 527 DDGRIIEAQHQQAWDVAILLEBFVEMWGDDEISLDLFOQMIEAGA---ESLTFSLIPP 583  
QY 485 LLNGQVRVMDGVSDHSCQSLQLTNVAATSGQYRGQTSPTHPNE-----K 512  
DB 584 AL-----DQVFVGN--MOLSRMYGTSCTFVLGANDGVLPARPDENGVLSDDRWLK 633  
QY 513 FYGVTVFKALKLGQEGKVPLOSAHLVYNTVEKVRIMESYERLDTF---LYFSY--SHLV 567  
DB 634 TIGV-----ELSSGGRRLLEHFLI-----YMAFSSPSDRLYVSYPIADAE 675  
QY 568 CHTALEEVOAERKDDSHPVHVNDNCILNAETLVCKEPPAYTFRYSAILYLNGDFDGNF 627  
DB 676 GXTLLPSMIVKRLLEELFPHHKERLLTN-----EPEQVS--DEEQMLYVNVKSVQSP 725  
QY 628 YTELDKATVTAQVOPQCGRVGFSSGTENPHGVKAVTRGQRCALALWFTLLDPRHSERDR 687  
DB 726 TASQL-----RLMTREYDISDVWVWSTVNVLMSEQDR 756  
QY 688 VQADDLVKMLFSPPEMDLSQOPLDAQGPPPEAQESLSGSES 730  
DB 757 LQSKKLFSSLFRRNE--VKQLERSVSRQLYGRIQGSVSRMET 797

## RESULT 15

MYSA\_CABEL STANDARD; PRT; 1969 AA.  
ID MYSA\_CABEL  
AC P12844; Q21440;  
DT 01-OCT-1989 (Rel. 12, Created)  
DT 01-OCT-1989 (Rel. 12, Last sequence update)  
DT 15-MAR-2004 (Rel. 43, Last annotation update)  
DE Myosin heavy chain A (YHC A).  
DE MIO-3 OR K12F2.1.  
OS Caenorhabditis elegans.  
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoides;  
OC Rhabditidae; Peloderinae; Caenorhabditis.  
OX NCBI\_TaxID=6239;



[illegible]

Search completed: July 18, 2004, 09:41:14  
Job time : 57 secs

This Page Blank (uspto)

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - nucleic search, using frame\_plus\_p2n model

Run on: July 18, 2004, 17:57:06 ; Search time 6546 Seconds  
(without alignments)  
3357.557 Million cell updates/sec

Title: US-10-045-815-4  
Perfect score: 736  
Sequence: 1 MAVRALKLLTLLAVAAAS.....PPEPAQESLSGSEKPKDEL 736

Scoring table:  
OLIGO  
Xgapop 60.0 , Xgapext 60.0  
Ygapop 60.0 , Ygapext 60.0  
Fgapop 6.0 , Fgapext 7.0  
Delop 6.0 , Delext 7.0

Searched: 27513289 seqs, 14931090276 residues

Word size: 1

Total number of hits satisfying chosen parameters: 55025477

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Command line parameters:

-MODEL=frame+ p2n.model -DEV=xlp  
-Q=/cgn2/1/USPRO.spool\_P/US10045815/runat\_14072004\_123119\_17251/app\_query.fasta\_1.903  
-DB=EST -QFMT=fastap -SUFFIX=oli.rst -MINMATCH=0.1 -LOOPCL=0 -LOOPEXT=0  
-UNITS=bits -START=1 -END=1 -MATRIX=oligo -TRANS=human40.cdi -LIST=45  
-DOCALLIGN=200 -THR SCORE=quality -THR MIN=1 -ALIGN=15 -MODE=LOCAL -OUTFMT=ptp  
-NORM=ext -HEAPSIZ=500 -MINLEN=0 -MAXLEN=2000000000  
-USER=US10045815@cgn.1.6425@runat\_14072004\_123119\_17251 -NCPU=5 -ICPU=3  
-NO\_WMAP -LARGEQUERY -NEG\_SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG  
-DEV TIMEOUT=120 -WARM TIMEOUT=30 -THREADS=1 -XGAPOP=60 -XGAPEXT=60 -FGAPOP=6  
-FGAPEXT=7 -YGAPOP=60 -YGAPEXT=60 -DELOP=6 -DELEXT=7

Database : EST:  
1: em\_estba:\*  
2: em\_esthum:\*  
3: em\_estin:\*  
4: em\_estmu:\*  
5: em\_estov:\*  
6: em\_estpl:\*  
7: em\_estro:\*  
8: em\_hrc:\*  
9: gb\_estl:\*  
10: gb\_est2:\*  
11: gb\_hrc:\*  
12: gb\_est3:\*  
13: gb\_est4:\*  
14: gb\_est5:\*  
15: em\_estfun:\*  
16: em\_eston:\*  
17: em\_gss\_hum:\*  
18: em\_gss\_inv:\*  
19: em\_gss\_pln:\*  
20: em\_gss\_vrt:\*  
21: em\_gss\_fun:\*  
22: em\_gss\_man:\*  
23: em\_gss\_mus:\*  
24: em\_gss\_pro:\*  
25: em\_gss\_rod:\*  
26: em\_gss\_phg:\*  
27: em\_gss\_vrl:\*  
28: gb\_gse1:\*

29: gb\_gss2:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB	ID	Description
1	359	48.8	2583	11	BC004160	BC004160 Homo sapi
2	245	33.3	848	13	BUI77586	BUI77586 AGENCOURT
3	244	33.2	905	13	BUI47415	BUI47415 AGENCOURT
4	240	32.6	929	13	BQ677232	BQ677232 AGENCOURT
5	237	32.2	899	13	BUI19488	BUI19488 AGENCOURT
6	230	31.2	884	13	BUI190877	BUI190877 AGENCOURT
7	194	26.4	937	10	BF206567	BF206567 601870908
8	192	26.1	989	13	BQ678884	BQ678884 AGENCOURT
9	190	25.8	705	10	BE294207	BE294207 601172737
10	187	25.4	885	13	BQ680853	BQ680853 AGENCOURT
11	185	25.1	870	13	BQ676738	BQ676738 AGENCOURT
12	181	24.6	899	13	BUI148566	BUI148566 AGENCOURT
13	178	24.2	881	13	BUI543499	BUI543499 AGENCOURT
14	174	23.6	955	13	BQ678346	BQ678346 AGENCOURT
15	168	22.8	730	10	AM964170	AM964170 EST376243
16	165	22.4	675	13	EX109119	EX109119 BX109119
17	163	22.1	575	9	AM068699	AM068699 cn21h12.x
18	162	22.0	823	13	BUI59350	BUI59350 AGENCOURT
19	160	21.7	941	13	BQ679334	BQ679334 602628511
20	160	21.7	941	13	BUI38872	BUI38872 AGENCOURT
21	155	21.1	883	12	BQ681584	BQ681584 602628071
22	153	20.8	840	12	B1114871	B1114871 602861453
23	152	20.7	555	12	BQ981273	BQ981273 IL5-CN006
24	152	20.7	1108	12	BM473346	BM473346 AGENCOURT
25	151	20.5	1101	13	EX364036	EX364036 60364036
26	148	20.1	981	13	BUI39422	BUI39422 AGENCOURT
27	148	20.1	1201	13	EX334235	EX334235 60334235
28	147	20.0	592	14	CB369351	CB369351 T8EST2V98
29	147	20.0	633	10	AM966813	AM966813 EST378887
30	147	20.0	867	14	CF265106	CF265106 AGENCOURT
31	147	20.0	999	14	CA488534	CA488534 AGENCOURT
32	145	19.7	926	9	AL521774	AL521774 AL521774
33	144	19.6	594	10	BF343590	BF343590 602014425
34	143	19.4	668	10	BE298108	BE298108 601118250
35	143	19.4	1041	13	EX364035	EX364035 60364035
C 36	140	19.0	651	14	CB055054	CB055054 NISC_SM07
C 37	140	19.0	1201	13	EX337092	EX337092 BX337092
C 38	140	19.0	528	9	AU280069	AU280069 AU280069
39	138	18.8	957	12	BQ420284	BQ420284 602448302
40	137	18.6	766	12	BI856252	BI856252 603382985
41	133	18.1	766	12	BM802225	BM802225 AGENCOURT
42	133	18.1	959	12	CB055053	CB055053 NISC_SM07
C 43	132	17.9	686	14	AM838215	AM838215 QV2-LT005
C 44	128	17.4	434	10	AM838215	AM838215 QV2-LT005
45	128	17.4	777	12	BG120604	BG120604 602346739

ALIGNMENTS

RESULT 1	BC004160	Homo sapiens leucine proline-enriched proteoglycan (leprecan) 1, mRNA (CDNA clone IMAGE:2824480), containing frame-shift errors.	2583 bp	mRNA	linear	HTC 19-NOV-2003
LOCUS	BC004160	Homo sapiens leucine proline-enriched proteoglycan (leprecan) 1, mRNA (CDNA clone IMAGE:2824480), containing frame-shift errors.	2583 bp	mRNA	linear	HTC 19-NOV-2003
DEFINITION	BC004160	Homo sapiens leucine proline-enriched proteoglycan (leprecan) 1, mRNA (CDNA clone IMAGE:2824480), containing frame-shift errors.	2583 bp	mRNA	linear	HTC 19-NOV-2003
ACCESSION	BC004160	Homo sapiens leucine proline-enriched proteoglycan (leprecan) 1, mRNA (CDNA clone IMAGE:2824480), containing frame-shift errors.	2583 bp	mRNA	linear	HTC 19-NOV-2003
VERSION	BC004160.2	Homo sapiens leucine proline-enriched proteoglycan (leprecan) 1, mRNA (CDNA clone IMAGE:2824480), containing frame-shift errors.	2583 bp	mRNA	linear	HTC 19-NOV-2003
KEYWORDS	HTC.	Homo sapiens (human)				
SOURCE	HTC.	Homo sapiens (human)				
ORGANISM	HTC.	Homo sapiens				
REFERENCE	1	(bases 1 to 2583)				



```
QY 380 yrAspValPheGlyIleProPheValAspProAspSerThrProGluGluValIleP 400
Db 1169 ATGATGTTTTTGGAAATCCCTTTGTGGATCCGATTCATGAGCTCCAGAGAAAGTGAATC 1228
QY 400 roLysArgLeuGlnGluLysGlnLysSerGluArgGluThrAlaValArgIleSerGlnG 420
Db 1229 CCAAGAGATTGCAAGAGAAACAGAGATCAGAACCGGAAACAGCGGTACGCATCTCCACG 1288
QY 420 lulleGlyAsnLeuMetLysGluIleGluThrLeuValGluGluLysThrLysGluSerL 440
Db 1289 AGATTGGGAACCTTATGAAGAAATCGAGACCTTGTGGAAGAGAGAACCAAGAGTCA 1348
QY 440 euAspValSerArgLeuThrArgGluGlyGlyProLeuLeuTyrGluGluIleSerLeu 460
Db 1349 TGGATGTGACAGACTGACCCGGAGGTGGCCCTGCTGTATGAAGCATCAGTCTCA 1408
QY 460 hrYerAsnSerLysLeuLeuAsnGlyTyr-GlnArgValValMetAspGlyValIleSer 479
Db 1409 CCATGAATCTCAAACTCCTGAATGGTTC-CCAGCGGTGGTGTGATGACCGCGTAATCTCT 1467
QY 480 AspHisGluCysGlnGluLeuGlnArgLeuThrAsnValAlaAlaThrSerGlyAspGly 499
Db 1468 GACCACAGGTGTCCAGAGCTGCAGAGACTACCAATGTGCAGCAACCTCAGAGATGGC 1527
QY 500 TyrArgGlyGlnThrSerProHisThrProAsnGluLysPheTyrGlyValThrValPhe 519
Db 1528 TACCGGGTTCAGACCTCCCCACATACCTCCCAATGAAAGTTCTATGGTGTCACTGTCTTC 1587
QY 520 LysAlaLeuLysLeuGlyGlnGluLysValProLeuGlnSerAlaHisLeuTyrTyr 539
Db 1588 AAAGCCCTCAAGCTGGGGCAAGAGCAAAAGTTCTCTCGAGAGTGCACCTTACTACTAC 1647
QY 540 AsnValThrGluLysValArgArgIleMetGluSerTyrPheArgLeuAspThrProLeu 559
Db 1648 AACGTGACGAGAGAGTGGCGGCATCATGAGTCTCTTCCGCTGGATAGCCCTC 1707
QY 560 TyrPheSerTyrSerHisLeuValCysArgThrAlaIleGluGluValGlnAlaGluArg 579
Db 1708 TACTTTTCTTCTACTCTCATCTGTTGCGCACTGCCATCGCAAGAGGTCTCCAGGAGAG 1767
QY 580 LysAspAspSerHisProValHisValAspAsnCysIleLeuAsnAlaGluThrLeuVal 599
Db 1768 AAGGATGATGTATCCAGTCCAGTGGCAACTGCAATCTGTAATGCCGAGACCTCGTG 1827
QY 600 CysValLysGluProProAlaTyrThrPheArgAspTyrSerAlaIleLeuTyrLeuAsn 619
Db 1828 TGTGTCAAAGAGCCCCAGCCTACACCTTCGCGACTACAGCGCCATCTTACTAAAT 1887
QY 620 GlyAspPheAspGlyGlyAsnPheTyrPheThrGluLeuAspAlaLysThrValThAla 639
Db 1888 GGGGACTTCGATGGCGGAAACTTTTATTCTACTGAATGCGCAAGACCGTGACGGCA 1947
QY 640 GluValGlnProGlnCysGlyArgAlaValGlyPheSerSerGlyThrGluAsnProHis 659
Db 1948 GAGTGCAGCCTCAGTGTGGAAGACCGTGGGATTCTCTTCAGGCACTGAAACCCACAT 2007
QY 660 GlyValLysAlaValThrArgGlyGlnArgCysAlaIleAlaLeuThrPheThrLeuAsp 679
Db 2008 GGAGTGAAGGTGTCAACAGGGGCGCGTGTGCCATCGCCCTGTGTTCACCCCTGGAC 2067
QY 680 ProArgHisSerGluArgAspArgValGlnAlaAspAspLeuValLysMetLeuPheSer 699
Db 2068 CCTGCACACAGCGGACAGGGTGCAGGCAGATGACCTGTGTAAGATGCTCTTCAGC 2127
QY 700 ProGluGluMetAspLeuSerGlnGluGlnProLeuAspAlaGlnGlnGlyProProGlu 719
Db 2128 CCAGAGAGATGGACCTCTCCAGAGAGCAGCCCCCTGGATGCCAGCAGGGGCCCCCGGAA 2187
QY 720 ProAlaGlnGluSerLeuSerGlySerGluSerLysProLysAspGluLeu 736
Db 2188 CCTGCACAAAGAGTCTCTCTCAGGAGTGAAATCGAAGCCCAAGGATGAGCTA 2238
```

## RESULT 2

```
BUI77586
LOCUS BUI77586
DEFINITION AGENCOURT_7954111 NIH_MGC_112 Homo sapiens cDNA clone IMAGE:6109712
5', mRNA sequence.
ACCESSION BUI77586
VERSION BUI77586.1 GI:22691570
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 848)
NIH-MGC http://mgc.nci.nih.gov/.
National Institutes of Health, Mammalian Gene Collection (MGC)
Unpublished (1999)
Contact: Robert Strausberg, Ph.D.
Email: cgapbs-remail.nih.gov
Tissue Procurement: DCTD/DP
cDNA Library Preparation: Rubin Laboratory
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Agencourt Bioscience Corporation
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: LLCM2J56 row: a column: 09
High quality sequence stop: 635.
FEATURES
Location/Qualifiers
1..848
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="IMAGE:6109712"
/lab_host="DH10B (phage-resistant)"
/tissue_type="melanotic melanoma, cell line"
/clone_lib="NIH MGC 112"
/note="Organ: skin; Vector: pOTB7; Site 1: XhoI; Site 2:
EcoRI; cDNA made by oligo-dT priming. Directionally cloned
into EcoRI/XhoI sites using the following 5' adaptor:
GCCACGAG(G). Library constructed by Ling Hong in the
laboratory of Gerald M. Rubin (University of California,
Berkeley) using ZAP-cDNA synthesis kit (Stratagene) and
Superscript II RT (Life Technologies). Note: this is a
NIH_MGC Library."
```

## ORIGIN

```
Alignment Scores:
Pred. No.: 8..61e-221 Length: 848
Score: 245.00 Matches: 245
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 33.29% Indels: 0
DB: 13 Gaps: 0

US-10-045-815-4 (1-736) x BUI77586 (1-848)
QY 471 ArgValValMetAspGlyValIleSerAspHisGluCysGlnGluLeuGlnArgLeuThr 490
Db 2 CGGGTGGTATCGAGCGGCGTAACTCTGACACAGAGTGTCCAGAGCTGCAGAGATGACC 61
QY 491 AsnValAlaAlaThrSerGlyAspGlyTyrArgGlyGlnThrSerProHisThrProAsn 510
Db 62 AATGTGGCAGCAACCTCAGGAGATGGTACC GGGTTCAGACCTCCCACTACTCCCAAT 121
QY 511 GluLysPheTyrGlyValThrValPheLysAlaLeuLysLeuGlyGlnGluLysVal 530
Db 122 GAAAGATTCTATGGTGTCTACTGTCTTCAAAGCCCTCAAGCTGGGCAAGAGCAAGT 181
QY 531 ProLeuGlnSerAlaHisLeuTyrTyrAsnValThrGluLysValArgArgIleMetGlu 550
Db 182 CTTCTGCAGAGTGCACCTGTACTACACGTGCAGGAGAGAGTGGCGGCATCATGGAG 241
QY 551 SerTyrPheArgLeuAspThrProLeuTyrPheSerTyrSerHisLeuValCysArgThr 570
```

|||||

242 TCCTACTTCCCGCTGGATACGCCCTCTACTTTCTTCTACTCTCATCTGGTGGCGCACT 301

Qy

571 AlalleGluValGlnAlaGluArgLysAspSerHisProValHisValAspAsn 590

Db

302 GCCATCGAAGAGTCCAGGACAGAGAGAGGATGATAGTCATCCAGTCCAGTGGACAAC 361

Qy

591 CysIleLeuAsnAlaGluThrLeuValCysValLysGluProProAlaThrPheArg 610

Db

362 TGCATCTCTGAATGCCGAGACCTCGTGTGTGTCGAAGAGGCCCGCAGCTACACTCCGC 421

Qy

611 AspTyrSerAlaIleLeuTyrLeuAsnGlyAspPheAspGlyGlyAsnPheTyrPheThr 630

Db

422 GACTACAGCGCATCTTACTTAATGGGACTTCGATGGGAAACTTTTATTTCACT 481

Qy

631 GluLeuAspAlaLysThrValThAlaGluValGlnProGlnCysGlyArgAlaValGly 650

Db

482 GAACCTGGATGCAAGACCGTACGCGCAGAGGTGCGCTCAGTGTGGAAGAGCGGTGGGA 541

Qy

651 PheSerSerGlyThrGluAsnProHisGlyValLysAlaValThrArgGlyGlnArgCys 670

Db

542 TTCCTTCAGGCACTGAAACCCACATGGAGTGAAGCTGTTCACCGGGGCGAGCGCTGT 601

Qy

671 AlalleAlaLeuThrPheThrLeuAspProArgHisSerGluArgAspArgValGlnAla 690

Db

602 GCCATCGCCCTGTGTTCCACCTCGACCTCGACACAGCGAGCGGGACAGGGTGCAGGCA 661

Qy

691 AspAspLeuValLysMetLeuPheSerProGluGluMetAspLeuSerGlnGluPro 710

Db

662 GATGACCTGGTGAAGATGCTCTTCAGCCCAAGAGATGAGCTCTCCCGAGGACGCCCC 721

Qy

711 LeuAspAlaGlnGln 715

Db

722 CTGGATGCCCGACGAG 736

RESULT 3

LOCUS BUI47415

DEFINITION AGENCOURT\_8071055 NIH\_MGC\_112 Homo sapiens cDNA clone IMAGE:6090065

5', mRNA sequence.

ACCESSION BUI47415

VERSION BUI47415.1 GI:22660947

KEYWORDS EST.

SOURCE Homo sapiens (human)

ORGANISM Homo sapiens

REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

AUTHORS 1 (bases 1 to 905)

TITLE NIH-MGC <http://mgi.nci.nih.gov/>.

JOURNAL National Institutes of Health, Mammalian Gene Collection (MGC)

COMMENT Contact: Robert Strausberg, Ph.D.  
Email: cgapbs-remail.nih.gov  
Tissue Procurement: DCTD/DTF  
CDNA Library Preparation: Rubin Laboratory  
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)  
DNA Sequencing by: Agencourt Bioscience Corporation  
Clone Distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LNL at:  
<http://image.llnl.gov>  
Plate: LLCM2330 row: n column: 18  
High quality sequence stop: 635.

FEATURES

source Location/Qualifiers

1..905

/organism="Homo sapiens"

/mol\_type="mRNA"

/db\_xref="taxon:9606"

/clone="IMAGE:6090065"

/tissue\_type="melanotic melanoma, cell line"

/lab\_host="DH10B (phage-resistant)"

/clone\_lib="NIH\_MGC\_112"

/notes="Organ: skin; Vector: pOTB7; Site\_1: XhoI; Site\_2: EcoRI; cDNA made by oligo-dT priming. Directionally cloned

into EcoRI/XhoI sites using the following 5' adaptor:  
GCCACGAG(G). Library constructed by Ling Hong in the  
Laboratory of Gerald M. Rubin (University of California,  
Berkeley) using ZAP-cDNA synthesis kit (Stratagene) and  
Superscript II RT (Life Technologies). Note: this is a  
NIH\_MGC Library."

## ORIGIN

Alignment Scores:

Pred. No.: 8,186-220 Length: 905

Score: 244.00 Matches: 244

Percent Similarity: 100.00% Conservative: 0

Best Local Similarity: 100.00% Mismatches: 0

Query Match: 33.15% Indels: 0

DB: 13 Gaps: 0

US-10-045-815-4 (1-736) x BUI47415 (1-905)

Qy 471 AsgValValMetAspGlyValIleSerAspHisGluCysGlnGluLeuGlnArgLeuThr 490

Db 2 CGGGTGGTGTGGAGCGGGTATCTCTGACCCACGAGTGTGAGGAGCTGCAGAGACTGACC 61

Qy 491 AsnValAlaAlaThrSerGlyAspGlyTyrArgGlyGlnThrSerProHisThrProAsn 510

Db 62 AATGTGGCAGCAACCTCAGAGAGTGGCTACCGGGGTGACAGCTCCCCACATATCCCAAT 121

Qy 511 GluLysPheTyrGlyValThrValPheLysAlaLeuLysLeuGlyGlnGluLysVal 530

Db 122 GAAAAGTTCTATGGTGTCTCTGTTCAAAGCCCTCAAGCTGGGGCAGAAGGCAAGTT 181

Qy 531 ProLeuGlnSerAlaHisLeuTyrTyrAsnValThrGluLysValArgArgIleMetGlu 550

Db 182 CCTCTGCAGAGTGGCCACCTGTACTTACAACTGACGAGAGAGGTGGCGGCATCATGGAG 241

Qy 551 SerTyrPheArgLeuAspThrProLeuTyrPheSerTyrSerHisLeuValCysArgThr 570

Db 242 TCCTACTTCCGCTGGATACGCCCTCTACTTTTCTACTCTCATCTGTGTGGCGCACT 301

Qy 571 AlalleGluValGlnAlaGluArgLysAspSerHisProValHisValAspAsn 590

Db 302 GCCATCGAAGAGTCCAGGACAGAGAGGATGATAGTCATCCAGTCCAGTGGACAAC 361

Qy 591 CysIleLeuAsnAlaGluThrLeuValCysValLysGluProProAlaThrPheArg 610

Db 362 TGCATCTCTGAATGCCGAGACCTCGTGTGTGTCGAAGAGGCCCGCAGCTACACTCCGC 421

Qy 611 AspTyrSerAlaIleLeuTyrLeuAsnGlyAspPheAspGlyGlyAsnPheTyrPheThr 630

Db 422 GACTACAGCGCATCTTACTTAATGGGACTTCGATGGGAAACTTTTATTTCACT 481

Qy 631 GluLeuAspAlaLysThrValThAlaGluValGlnProGlnCysGlyArgAlaValGly 650

Db 482 GAACCTGGATGCAAGACCGTACGCGCAGAGGTGCGCTCAGTGTGGAAGAGCGGTGGGA 541

Qy 651 PheSerSerGlyThrGluAsnProHisGlyValLysAlaValThrArgGlyGlnArgCys 670

Db 542 TTCCTTCAGGCACTGAAACCCACATGGAGTGAAGCTGTTCACCGGGGCGAGCGCTGT 601

Qy 671 AlalleAlaLeuThrPheThrLeuAspProArgHisSerGluArgAspArgValGlnAla 690

Db 602 GCCATCGCCCTGTGTTCCACCTCGACCTCGACACAGCGAGCGGGACAGGGTGCAGGCA 661

Qy 691 AspAspLeuValLysMetLeuPheSerProGluGluMetAspLeuSerGlnGluPro 710

Db 662 GATGACCTGGTGAAGATGCTCTTCAGCCCAAGAGATGAGCTCTCCCGAGGACGCCCC 721

Qy 711 LeuAspAlaGln 714

Db 722 CTGGATGCCCGAG 733

## RESULT 4

BQ677232

LOCUS

BQ677232 929 bp mRNA linear EST 15-JUL-2002

```

DEFINITION AGENCOURT 8494737 NIH_MGC_112 Homo sapiens cDNA clone IMAGE:6294988
5', mRNA sequence.
ACCESSION B0677232
VERSION B0677232.1 GI:21789911
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 929)
AUTHORS NIH-MGC http://mgc.nci.nih.gov/
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgapbs-x@mail.nih.gov
Tissue Procurement: DCTD/DTF
cDNA Library Preparation: Rubin Laboratory
DNA Sequencing by: The I.M.A.G.E. Consortium (LLNL)
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: LCM2500 row: i column: 05
High quality sequence start: 5
High quality sequence stop: 638.
FEATURES
    source
        1..929
            /organism="Homo sapiens"
            /mol_type="mRNA"
            /db_xref="taxon:9606"
            /clone="IMAGE:6294988"
            /tissue_type="melanotic melanoma, cell line"
            /lab_host="DH10B (phage-resistant)"
            /clone_lib="NIH_MGC_112"
            /notes="Organ: skin; Vector: pOTB7; Site 1: XhoI; Site 2:
            into EcoRI; cDNA made by oligo-dT priming. Directionally cloned
            into EcoRI/XhoI sites using the following 5' adaptor:
            GGACACAG(G). Library constructed by Ling Hong in the
            Laboratory of Gerald M. Rubin (University of California,
            Berkeley) using ZAP-cDNA synthesis kit (Stratagene) and
            Superscript II RT (Life Technologies). Note: this is a
            NIH_MGC Library."
ORIGIN
Alignment Scores:
Pred. No.: 5,12e-216 Length: 929
Score: 240.00 Matches: 240
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 32.61% Indels: 0
DB: 13 Gaps: 0
US-10-045-815-4 (1-736) x B0677232 (1-929)
QY 497 GlyAspGlyTyrArgGlyGlnThrSerProHisThrProAsnGluLysPheTyrGlyVal 516
Db 17 GGAGATGGCTACCGGGGTGAGACCTCCACATACCTCCCAATGAAAGTTCATGTGTGC 76
QY 517 ThrValPheLysAlaLeuLysLeuGlyGlnGluLysValProLeuGlnSerAlaHis 536
Db 77 ACTGTCTTCAAGCCCTCAAGCTGGGGCAAGAGGCAAGATTCCTCTGCAGAGTGGCCAC 136
QY 537 LeuTyrTyrAsnValThrGluLysValArgArgGlyMetGluSerTyrPheArgLeuAsp 556
Db 137 CTGTACTACACGTGACGGAGAGGTGGCGCATCATGAGTCTTCTTCCGCTGGAT 196
QY 557 ThrProLeuTyrPheSerTyrSerHisLeuValCysArgThrAlaIleGluGluValGln 576
Db 197 ACGCCCTCTACTTTCTCTACTCTCTATCTGGTGGCGCACTGCCATCGAAGAGGTCCAG 256
QY 577 AlaGluArgLysAspSerHisProValHisValAspAsnCysIleLeuAsnAlaGlu 596
Db 257 GCAGAGGAGGAGGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 316

```

```

QY 597 ThrLeuValCysValLysGluProProAlaTyrThrPheArgAspTyrSerAlaIleLeu 616
Db 317 ACCCTCGTGTGTCTCAAGAGCCCCCAGCCTACCTTCCGCGACTAGAGGCCATCTT 376
QY 617 TyrLeuAsnGlyAspPheAspGlyGlyAsnPheTyrPheThrGluLeuAspAlaLysThr 636
Db 377 TACCTAAATGGGACTTCGATGGCGGAACTTTTATTTCACTGAATCGATGCCAAGACC 436
QY 637 ValThrAlaGluValGlnProGlnCysGlyArgAlaValGlyPheSerSerGlyThrGlu 656
Db 437 GTGACGGCAGAGGTGCAGCTCAGTGTGGAAGAGCCGTGGGATTTCTTTCAGGCTGAA 496
QY 657 AsnProHisGlyValLysAlaValThrArgGlyGlnArgCysAlaIleAlaLeuTyrPhe 676
Db 497 AACCCACATGGAGTGAAGCTGTACACAGGGGGCAGCGCTGTGCCATCGCCTGTGGTTC 556
QY 677 ThrLeuAspProArgHisSerGluArgAspArgValGlnAlaAspAspLeuValLysMet 596
Db 557 ACCCTGGACCTTCGACACAGCGAGCGGAGGTGCGAGGAGATGACCTGGTGAAGATG 616
QY 697 LeuPheSerProGluGluMetAspLeuSerGlnGluGlnProLeuAspAlaGlnGlnGly 716
Db 617 CTCCTCAGCCCAAGAGATGGACCTCTCCAGGAGCAGCCCTGGATGCCAGCAGGCG 676
QY 717 ProProGluProAlaGlnGluSerLeuSerGlySerGluSerLysProLysAspGluLeu 736
Db 677 CCCCCGAACTGCACAAAGAGTCTCTCTCAGCAGTGAATCGAAGCCCAAGATGAGCTA 736

```

```

RESULT 5
BUI79488
LOCUS BUI79488
DEFINITION BUI79488 899 bp mRNA linear EST 04-SEP-2002
5', mRNA sequence.
ACCESSION BUI79488
VERSION BUI79488.1 GI:22693472
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

```

```

REFERENCE 1 (bases 1 to 899)
AUTHORS NIH-MGC http://mgc.nci.nih.gov/.
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgapbs-remail.nih.gov
Tissue Procurement: DCTD/DTF
cDNA Library Preparation: Rubin Laboratory
DNA Sequencing by: The I.M.A.G.E. Consortium (LLNL)
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: LCM2291 row: i column: 24
High quality sequence stop: 627.
FEATURES
    Location/Qualifiers
        1..899
            /organism="Homo sapiens"
            /mol_type="mRNA"
            /db_xref="taxon:9606"
            /clone="IMAGE:6074975"
            /tissue_type="melanotic melanoma, cell line"
            /lab_host="DH10B (phage-resistant)"
            /clone_lib="NIH_MGC_112"
            /note="Organ: skin; Vector: pOTB7; Site 1: XhoI; Site 2:
            EcoRI; cDNA made by oligo-dT priming. Directionally cloned
            into EcoRI/XhoI sites using the following 5' adaptor:
            GGACACAG(G). Library constructed by Ling Hong in the
            Laboratory of Gerald M. Rubin (University of California,
            Berkeley) using ZAP-cDNA synthesis kit (Stratagene) and
            Superscript II RT (Life Technologies). Note: this is a
            NIH_MGC Library."

```





```

422 ACGGCAGAGGTGCAGCCTCAGTGTGGAAGAGCGGTGGGATTCTTTCAGGCACCTGAAAC 481
423 ValAlaAlaThrSerGlyAspGlyTyrArgGlyGlnThrSerProHisThrProAsnGlu 511
424 GTGGCAGCAACTCAGGAGATGGCTACCGGGGTGAGACCTCCACATACCTCCCAATGAA 63
425 LysPheTyrGlyValThrValPheLysAlaLeuLysLeuGlyGlnGluGlyLysValPro 531
426 AAGTTCTATGTGTGCTACTGTCTTCAAGAGCCCTCAAGCTGGGGCAGGAAGGCAAAATTCCT 123
427 LeuGlnSerAlaHisLeuTyrTyrAsnValThrGluLysValArgArgIleMetGluSer 551
428 CTGAGAGTGGCCACCTGTACTACAGTACGAGAGAGAGGTGGCGGCATCATCGAGATCC 183
429 TyrPheArgLeuAspThrProLeuTyrPheSerTyrSerHisLeuValCysArgThrAla 571
430 TACTTCGGCTGGATAGCCCTCTACTTTTCTTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 243
431 IleGluGluValGlnAlaGluArgLysAspSerHisProValHisValAspAsnCys 591
432 ATCGAAGAGGTCCAGGAGAGAGAGATATAGTATCATCGTCCAGTCCAGTCCAGTCCAGTCC 303
433 IleLeuAsnAlaGluThrLeuValCysValLysGluProProAlaTyrThrPheArgAsp 611
434 ATCTGAATGCGAGAGCCTCTGTGTGTCAAGAGCCCCCAGCCTACACCTTCCGGGAC 363
435 TyrSerAlaIleLeuTyrLeuAsnGlyAspPheAspGlyGlyAsnPheTyrPheThrGlu 631
436 TACAGCGCCATCTTACTTAATGGGACTTCGATGGCGGAACTTTTATTTCCTGAA 423
437 LeuAspAlaLysThrValThrAlaGluValGlnProGlnCysGlyArgAlaValGlyPhe 651
438 CTGATGTCAGAGACCGTGACGCGAGAGGTGAGCTCAGTGTGGAGAGCGCTGGGATTC 483
439 SerSerGlyThrGluAsnProHisGlyValLysAlaValThrArgGlyGlnArgCysAla 671
440 TCTTCAGGCACCTGAAACCCACATGGAGTGAAGCTGTCAACAGGGGGGAGCGGTGTGCC 543
441 IleAlaLeuTyrPheThrLeuAspProArgHisSerGluArg 685
442 ATCGCCCTGTGGTTCACCTGGACCCCTGACACAGCAGCGAGCGG 585

RESULT 8
LOCUS      BQ678884
DEFINITION AGENCOURT_8497890 NIH_MGC_112 Homo sapiens cDNA clone IMAGE:6257808
5', mRNA sequence.
ACCESSION  BQ678884
VERSION     BQ678884.1
KEYWORDS    GI:21791563
SOURCE      Homo sapiens (human)
ORGANISM    Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE  1 (bases 1 to 989)
AUTHORS    NIH-MGC http://mgi.nci.nih.gov/.
TITLE      National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL    Unpublished (1999)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgaabs-remail.nih.gov
            Tissue Procurement: DCTD/DTF
            cDNA Library Preparation: Ruben Laboratory
            DNA Sequencing by: The I.M.A.G.E. Consortium (LLNL)
            Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: LLCM2412 row: 1 column: 01
            High quality sequence start: 46
            High quality sequence stop: 404.

FEATURES             Location/Qualifiers
     source            1..989
                        /organism="Homo sapiens"
                        /mol_type="mRNA"
                        /db_xref="taxon:9606"

Alignment Scores:
Pred. No.:          1.7e-172
Score:              194.00
Percent Similarity: 100.00%
Best Local Similarity: 100.00%
Query Match:       26.36%
Indels:             0
Gaps:               0
DB:                 0

US-10-045-815-4 (1-736) x BF206567 (1-937)

```

```

QY 492 ValAlaAlaThrSerGlyAspGlyTyrArgGlyGlnThrSerProHisThrProAsnGlu 511
DB 4 GTGGCAGCAACTCAGGAGATGGCTACCGGGGTGAGACCTCCACATACCTCCCAATGAA 63
QY 512 LysPheTyrGlyValThrValPheLysAlaLeuLysLeuGlyGlnGluGlyLysValPro 531
DB 64 AAGTTCTATGTGTGCTACTGTCTTCAAGAGCCCTCAAGCTGGGGCAGGAAGGCAAAATTCCT 123
QY 532 LeuGlnSerAlaHisLeuTyrTyrAsnValThrGluLysValArgArgIleMetGluSer 551
DB 124 CTGAGAGTGGCCACCTGTACTACAGTACGAGAGAGAGGTGGCGGCATCATCGAGATCC 183
QY 552 TyrPheArgLeuAspThrProLeuTyrPheSerTyrSerHisLeuValCysArgThrAla 571
DB 184 TACTTCGGCTGGATAGCCCTCTACTTTTCTTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 243
QY 572 IleGluGluValGlnAlaGluArgLysAspSerHisProValHisValAspAsnCys 591
DB 244 ATCGAAGAGGTCCAGGAGAGAGAGATATAGTATCATCGTCCAGTCCAGTCCAGTCCAGTCC 303
QY 592 IleLeuAsnAlaGluThrLeuValCysValLysGluProProAlaTyrThrPheArgAsp 611
DB 304 ATCTGAATGCGAGAGCCTCTGTGTGTCAAGAGCCCCCAGCCTACACCTTCCGGGAC 363
QY 612 TyrSerAlaIleLeuTyrLeuAsnGlyAspPheAspGlyGlyAsnPheTyrPheThrGlu 631
DB 364 TACAGCGCCATCTTACTTAATGGGACTTCGATGGCGGAACTTTTATTTCCTGAA 423
QY 632 LeuAspAlaLysThrValThrAlaGluValGlnProGlnCysGlyArgAlaValGlyPhe 651
DB 424 CTGATGTCAGAGACCGTGACGCGAGAGGTGAGCTCAGTGTGGAGAGCGCTGGGATTC 483
QY 652 SerSerGlyThrGluAsnProHisGlyValLysAlaValThrArgGlyGlnArgCysAla 671
DB 484 TCTTCAGGCACCTGAAACCCACATGGAGTGAAGCTGTCAACAGGGGGGAGCGGTGTGCC 543
QY 672 IleAlaLeuTyrPheThrLeuAspProArgHisSerGluArg 685
DB 544 ATCGCCCTGTGGTTCACCTGGACCCCTGACACAGCAGCGAGCGG 585

RESULT 8
LOCUS      BQ678884
DEFINITION AGENCOURT_8497890 NIH_MGC_112 Homo sapiens cDNA clone IMAGE:6257808
5', mRNA sequence.
ACCESSION  BQ678884
VERSION     BQ678884.1
KEYWORDS    GI:21791563
SOURCE      Homo sapiens (human)
ORGANISM    Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE  1 (bases 1 to 989)
AUTHORS    NIH-MGC http://mgi.nci.nih.gov/.
TITLE      National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL    Unpublished (1999)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgaabs-remail.nih.gov
            Tissue Procurement: DCTD/DTF
            cDNA Library Preparation: Ruben Laboratory
            DNA Sequencing by: The I.M.A.G.E. Consortium (LLNL)
            Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: LLCM2412 row: 1 column: 01
            High quality sequence start: 46
            High quality sequence stop: 404.

FEATURES             Location/Qualifiers
     source            1..989
                        /organism="Homo sapiens"
                        /mol_type="mRNA"
                        /db_xref="taxon:9606"

```

/clones="IMAGE:6257808"  
 /tissue\_type="melanotic melanoma, cell line"  
 /lab\_host="DH10B (phage-resistant)"  
 /clone\_lib="NIH\_MGC\_112"  
 /note="Organ: skin; Vector: pOTB7; Site 1: XhoI; Site 2:  
 EcoRI; cDNA made by oligo-dt priming. Directionally cloned  
 into EcoRI/XhoI sites using the following 5' adaptor:  
 GGACAGAG(G). Library constructed by Ling Hong in the  
 laboratory of Gerald M. Rubin (University of California,  
 Berkeley) using ZAP-cDNA synthesis kit (Stratagene) and  
 Superscript II RT (Life Technologies). Note: this is a  
 NIH\_MGC Library."

## ORIGIN

Alignment Scores:  
 Pred. No.: 1,41e-170 Length: 989  
 Score: 192.00 Matches: 192  
 Percent Similarity: 100.00% Conservative: 0  
 Best Local Similarity: 100.00% Mismatches: 0  
 Query Match: 26.09% Indels: 0  
 DB: 13 Gaps: 0

US-10-045-815-4 (1-736) x B0678884 (1-989)

Qy 470 GlnArgValValMetAspGlyValIleSerAspHisGluCysGlnGluLeuGlnArgLeu 489  
 Db 12 CAGCGGGTGGTGATGGAGCGGGTAACTCTGACCAGGTGTGAGGAGTGCAGAGACTG 71  
 Qy 490 ThrAsnValAlaThrSerGlyAspGlyTyrArgGlyGlnThrSerProHisThrPro 509  
 Db 72 ACCAATGTGGCAGCAACTCAGGAGATGGTACCGGGGTGAGACCTCCACATATCTCC 131  
 Qy 510 AsnGluLysPheTyrGlyValThrValPheLysAlaLeuLysLeuGlyGlnGluGlyLys 529  
 Db 132 AATGAAAGTTCTATGGTGTGCTACTGTCTCAAGCCCTCAAGTGGGGCAAGAGCANA 191  
 Qy 530 ValProLeuGlnSerAlaHisLeuTyrTyrAsnValThrGluLysValArgArgIleMet 549  
 Db 192 GTTCTCTGCGAGTGGCCACCTGTACTCAACGTGACGAGAGGFGCGCGGCATCATG 251  
 Qy 550 GluSerTyrPheArgLeuAspThrProLeuTyrPheSerTyrSerHisLeuValCysArg 569  
 Db 252 GAGTCTCTACTTCGCGCTGGATAGCCCTCTACTTTCTCTACTCTCTGTTGGTGGCG 311  
 Qy 570 ThrAlaIleGluValGlnAlaGluArgLysAspAspSerHisProValHisValAsp 589  
 Db 312 ACTGCCATCGAAGAGTCCAGGAGAGAGGATGATAGTATCATCCAGTCCACGTGGAC 371  
 Qy 590 AsnCysIleLeuAsnAlaGluThrLeuValCysValLysGluProProAlaTyrThrPhe 609  
 Db 372 AACTGTCATCTTGATGCGAGACCTCGTGTGTCTAAGAGCCCGCCAGCTACACTTC 431  
 Qy 610 ArgAspTyrSerAlaIleLeuTyrLeuAsnGlyAspPheAspGlyGlyAsnPheTyrPhe 629  
 Db 432 CGCGACTACAGCGCCATCTTTACTTAATGGGACTTCGATGGCGAAACTTTTATTC 491  
 Qy 630 ThrGluLeuAspAlaLysThrValThrAlaGluValGlnProGlnCysGlyArgAlaVal 649  
 Db 492 ACTGAATGTGATGCAAGACCGTGACGGCAGAGGTGCAGCCTCAGTGTGGAAGCCGTG 551  
 Qy 650 GlyPheSerSerGlyThrGluAsnProHisGlyVal 661  
 Db 552 GGATTCTCTTCAGGAGCTGAAACCCACATGGAGTG 587

RESULT 9  
 BE294207  
 LOCUS  
 DEFINITION 601172737F1 NIH\_MGC\_17 Homo sapiens cDNA clone IMAGE:3528273 5',  
 mRNA sequence.  
 BE294207  
 BE294207.1 GI:9177841  
 VERSION  
 KEYWORDS EST.  
 SOURCE Homo sapiens (human)

## ORGANISM

Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

## REFERENCE

1 (bases 1 to 705)

## AUTHORS

NIH-MGC http://mgc.ncbi.nlm.nih.gov/.

## TITLE

National Institutes of Health, Mammalian Gene Collection (MGC)

## JOURNAL

Unpublished (1999)

## COMMENT

Contact: Robert Strausberg, Ph.D.

Email: cgapbs-r@mail.nih.gov

Tissue Procurement: ATCC

cDNA Library Preparation: Ling Hong/Rubin Laboratory

cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)

DNA Sequencing by: Incyte Genomics, Inc.

Clone distribution: MGC clone distribution information can be

found through the I.M.A.G.E. Consortium/LLNL at: image.llnl.gov

Plate: LLCMI97 row: i column: 10

High quality sequence stop: 641.

Location/Qualifiers

1..705

/organism="Homo sapiens"

/mol\_type="mRNA"

/db\_xref="taxon:9606"

/clone="IMAGE:3528273"

/tissue\_type="rhabdomyosarcoma"

/lab\_host="DH10B (phage-resistant)"

/clone\_lib="NIH\_MGC\_17"

/note="Organ: muscle; Vector: pOTB7; Site 1: EcoRI;

Site 2: XhoI; cDNA made by oligo-dt priming.

Directionally cloned into EcoRI/XhoI sites using the

following 5' adaptor: GGACAGAG(G). Size-selected >500bp

for average insert size 1.8kb. Library constructed by

Ling Hong in the laboratory of Gerald M. Rubin (University

of California, Berkeley) using ZAP-cDNA synthesis kit

(Stratagene) and Superscript II RT (Life Technologies)."

## ORIGIN

Alignment Scores:  
 Pred. No.: 7,56e-169 Length: 705  
 Score: 190.00 Matches: 203  
 Percent Similarity: 99.51% Conservative: 0  
 Best Local Similarity: 99.51% Mismatches: 0  
 Query Match: 25.82% Indels: 1  
 DB: 10 Gaps: 0

US-10-045-815-4 (1-736) x BE294207 (1-705)

Qy 486 LeuGlnArgLeuThrAsnValAlaAlaThrSerGlyAspGlyTyrArgGlyGlnThrSer 505  
 Db 2 CTGCAGAGACTGACCAATGTGGCAGCAACCTCAGGAGATGGCTACCGGGGTGACACCTCC 61  
 Qy 506 ProHisThrProAsnGluLysPheTyrGlyValThrValPheLysAlaLeuLysLeuGly 525  
 Db 62 CCACATACCTCCCAATGAAAAGTTCTATGGTGTCTACTGTCTTCAAAGCCCTCAAGCTGGG 121  
 Qy 526 GlnGluGlyLysValProLeuGlnSerAlaHisLeuTyrTyrAsnValThrGluLysVal 545  
 Db 122 CAGAGAGCAAGTTCCTCTGCAGAGTGGCCACCTGTACTACACGTGACGAGAGAGTG 181  
 Qy 546 ArgArgIleMetGluSerTyrPheArgLeuAspThrProLeuTyrPheSerTyrSerHis 565  
 Db 182 CGCGCATCATGAGTCTCTTCCGCTGGATACGCGCCCTCTACTTTCCTACTCTCAT 241  
 Qy 566 LeuValCysArgThrAlaIleGluGluValGlnAlaGluArgLysAspAspSerHisPro 585  
 Db 242 CTGGTGTCCCGACTGCCATCGAAGGTCCAGCAGAGAGAGAGATGATGATCATCCA 301  
 Qy 586 ValHisValAspAsnCysIleLeuAsnAlaGluThrLeuValCysValLysGluProPro 605  
 Db 302 GTCCACGTGGACAACTGTCATCTGAATGCCGAGACCTCGTGTGTGTCAAGAGACCCCCA 361  
 Qy 606 AlaTyrThrPheArgAspTyrSerAlaIleLeuTyrLeuAsnGlyAspPheAspGlyGly 625  
 Db 362 GCCTACACCTTCCCGACTACAGGCCATCTTTTACCTAAATGGGACTTCGATGGCGGA 421

Qy	626	AsnPheTyrPheThrGluLeuAspAlaIySerValThrAlaGluValGlnProGlnCys	645
Db	422	AAC'TTTATTTCACTGAAC'TGGATGCCAAGACCGTGACGGCAGAGGTGCAGCTCAGTGT	481
Qy	646	GlvArgAlaValGlyPheSerSerGlyThrGluAsnProHisGlyValIyAlaValThr	665
Db	482	GGAAGAGCCGTCGGATCTCTTCAGGCAC'TGAAACCCACCATGGAGTGAAGGTGTCAAC	541
Qy	666	ArgGlyGlnArgCysAlaIleAlaLeu-TTpPheThrLeuAspProArghHisSerGluAr	695
Db	542	AGGGGGCAGCGCTGTGCATCGCCCTGTGGTTTCACCTCGACCTCGACACAGCGAGCG	601
Qy	685	gaspArgVal 688	
Db	602	GCACAGGGTG 611	
RESULT 10			
BQ680853			
LOCUS			
DEFINITION BQ680853 819221.3 NIH_MGC_112 Homo sapiens cDNA clone IMAGE:6260027			
5', mRNA sequence.			
ACCESSION BQ680853			
VERSION BQ680853.1 GI:21793532			
KEYWORDS EST.			
SOURCE Homo sapiens (human)			
ORGANISM Homo sapiens			
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;			
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.			
REFERENCE			
AUTHORS			
TITLE NH-MGC <a href="http://mgc.nci.nih.gov/">http://mgc.nci.nih.gov/</a> .			
JOURNAL National Institutes of Health, Mammalian Gene Collection (MGC)			
COMMENT Unpublished (1999)			
Contact: Robert Strausberg, Ph.D.			
Email: rcgapse-rc@mail.nih.gov			
Tissue Procurement: DCTD/BTP			
CDNA Library Preparation: Rubin Laboratory			
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)			
DNA Sequencing by: Agencourt Bioscience Corporation			
Clone Distribution: MGC clone distribution information can be			
found through the I.M.A.G.E. Consortium/LLNL at:			
<a href="http://image.llnl.gov">http://image.llnl.gov</a>			
Plate: LLCM2418 row: h column: 12			
High quality sequence stop: 534.			
FEATURES			
source			
1..885			
/organism="Homo sapiens"			
/mol_type="mRNA"			
/db_xref="taxon:9606"			
/clone="IMAGE:6260027"			
/tissue_type="melanotic melanoma, cell line"			
/lab_host="DH10B (phage-resistant)"			
/clone_lib="NIH_MGC_112"			
/note="Organ: skin; Vector: pOTB7; Site_1: XhoI; Site_2:			
EcoRI; CDNA made by oligo-dr priming. Directionally cloned			
into EcoRI/XhoI sites using the following 5' adaptor:			
GGCACGAG(G). Library constructed by Ling Hong in the			
laboratory of Gerald M. Rubin (University of California,			
Berkeley) using ZAP-cDNA synthesis kit (Stratagene) and			
Superscript II RT (Life Technologies). Note: this is a			
NIH_MGC Library."			
ORIGIN			
Alignment Scores:			
Pred. No.: 6,71e-166 Length: 885			
Score: 187.00 Matches: 187			
Percent Similarity: 100.00% Conservative: 0			
Best Local Similarity: 100.00% Mismatches: 0			
Query Match: 15.41% Indels: 0			
DB: 13 Gaps: 0			
US-10-045-815-4 (1-736) x BQ680853 (1-885)			

/tissue type="melanotic melanoma, cell line"  
/lab host="DH10B (phage-resistant)"  
/clone.lib="NIH\_MGC\_112"

/note="Organ: skin; Vector: pOTB7; Site: 1: XhoI; Site 2:  
EcoRI; cDNA made by oligo-dT priming. Directionally cloned  
into EcoRI/XhoI sites using the following 5' adaptor:  
GGCAGGAG(G). Library constructed by Ling Hong in the  
laboratory of Gerald M. Rubin (University of California,  
Berkeley) using ZAP-cDNA synthesis kit (Stratagene) and  
Superscript II RT (Life Technologies). Note: this is a  
NIH\_MGC Library."

## ORIGIN

## Alignment Scores:

Pred. No.: 5,13e-164 Length: 870  
Score: 185.00 Matches: 255  
Percent Similarity: 98.84% Conservative: 0  
Best Local Similarity: 98.84% Mismatches: 1  
Query Match: 25.14% Indels: 3  
DB: 13 Gaps: 0

US-10-045-815-4 (1-736) x B0676738 (1-870)

QY 412 GlnThrAlaValArgIleSerGlnGluIleGlyAsnLeuMetLysGluIleGluThrLeu 431  
DB 1 GAAACAGCGGACCGCATCTCCAGAGATGGGAACCTTATGAAGGAATCGAGACCCIT 60  
QY 432 ValGluGluThrLysGluSerLeuAspValSerArgLeuThrArgGluGlyGlyPro 451  
DB 61 GTGAAGAGAGACCAAGAGTCACTGATGTAGACAGACTGACCCGGAGGTGGCCCC 120  
QY 452 LeuLeuThrGluGlyIleSerLeuThrValMetAsnSerLysLeuLeuAsnGlyTyr-GlnAr 471  
DB 121 CTGCTGATGAAGGCATCAGTCTCACCATGAACCTCCAACTCTCGAATGCTC-CCAGCG 179  
QY 471 qValValMetAspGlyValIleSerAspHisGluCysGlnGluLeuGlnArgLeuThrAs 491  
DB 180 GTGGTGTATGACGGCGTATCTCTGACCAGAGTGTACAGAGCTGCAGAGATGACCAA 239  
QY 491 nValAlaAlaThrSerGlyAspGlyTyrArgGlyGlnThrSerProHisThrProAsnG 511  
DB 240 TGTGGCAGCAACCTCAGAGATGCTACCGGGGTACAGACCTCCCCACATACTCCCAATGA 299  
QY 511 uLysPheThrGlyValThrValPheLysAlaLeuLysLeuGlyGlnGluGlyValPr 531  
DB 300 AAAGTTCTATGGTGTCTACTGCTTCAAGCCCTCAGCTGGGGCAAGAGCGCAAGTTCC 359  
QY 531 oLeuGlnSerAlaHisLeuThrValThrGluLysValArgGlyMetGluSe 551  
DB 360 TCTCAGAGTCCCACTGCTACTACAACGTGACGAGAGGTGGCGGCATCATGGAGTC 419  
QY 551 rTyrPheArgLeuAspThrProLeuThrPheSerTyrSerHisLeuValCysArgThrAl 571  
DB 420 CTACTTCGCGCTGGATACGCCCTCTACTTTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 479  
QY 571 aileGluGluValGlnAlaGluArgLysAspAspSerHisProValHisValAspAsnCy 591  
DB 480 CATCGAAGAGGTCCAGGCAGAGAGGAAGGATGATGATCATCCAGTCCACGTGGACAACTG 539  
QY 591 sileLeuAsnAlaGluThrLeuValCysValLysGluProProAlaTyrThrPheArgAs 611  
DB 540 CATCCTGAATGCCAGACCCCTCGTGTGTGTAAGAGAGCCGCCACCTTACACCTTCGCGCA 599  
QY 611 pTyrSerAlaIleLeuThrValLeuAsnGlyAspPheAspGlyGlyAsnPheThrG 631  
DB 600 CTACAGCGCCATCTTTACCTTAATGGGACTTCGATGGCGGAACCTTTATTCTACTGA 659  
QY 631 uLeuAspAlaLysThrValThrAlaGluValGlnPro-GlnCysGlyArgAlaValGlyP 651  
DB 660 ACTGGATGCCAAGACCGTGAAGGAGAGGTGCAGCCNTCAGTGTGAAGAGCGCGTGGAT 719  
QY 651 heSerSerGlyThrGluAsnProHisGlyValLysAlaValThrArgGly 667

Db 720 TCTCTTCAGGCACCTGAAACCCACCATGGAGTGAAGGCTGTCCACGAGGGG 769

## RESULT 12

BUI48566

## LOCUS

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

COMMENT

1 (bases 1 to 899)

National Institutes of Health, Mammalian Gene Collection (MGC)

Unpublished (1999)

Contact: Robert Strausberg, Ph.D.

Email: cgabbs-remail.nih.gov

Tissue Procurement: DCTD/DTF

cDNA Library Preparation: Rubin Laboratory

DNA Sequencing by: The I.M.A.G.E. Consortium (LLNL)

Clone distribution: MGC clone distribution information can be

found through the I.M.A.G.E. Consortium/LLNL at:

http://image.llnl.gov

Plate: LLCW2568 row: n column: 15

High quality sequence stop: 724.

Location/Qualifiers

1..899

/organism="Homo sapiens"

/mol\_type="mRNA"

/db\_xref="taxon:9606"

/clone="IMAGE:6380366"

/tissue\_type="carcinoma, cell line"

/lab\_host="DH10B (phage-resistant)"

/clone\_lib="NIH\_MGC\_40"

/note="Organ: prostate; Vector: pOTB7; Site: 1: XhoI;

Site 2: EcoRI; cDNA made by oligo-dT priming.

Directionally cloned into EcoRI/XhoI sites using the

following 5' adaptor: GGCAGGAG(G). Library constructed by

Ling Hong in the laboratory of Gerald M. Rubin (University

of California, Berkeley) using ZAP-cDNA synthesis kit

(Stratagene) and Superscript II RT (Life Technologies).

Note: this is a NIH\_MGC Library."

## ORIGIN

## Alignment Scores:

Pred. No.: 3.24e-160 Length: 899  
Score: 181.00 Matches: 281  
Percent Similarity: 99.29% Conservative: 0  
Best Local Similarity: 99.29% Mismatches: 1  
Query Match: 24.59% Indels: 2  
DB: 13 Gaps: 0

US-10-045-815-4 (1-736) x BUI48566 (1-899)

QY 143 ArgSerProTyrAsnTyrLeuGlnValAlaTyrPheLysIleAsnLysLeuGluLysAla 162  
DB 3 CGAGACCCCTCAACTACTCTCAGTCCCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 62  
QY 163 ValAlaAlaAlaHisThrPhePheValGlyAsnProGluHisMetGluMetGlnGlnAsn 182  
DB 63 GTTGCTGCAGCACACACCTTCTCTGTTGGCAATCCTGAGCACATGGAAATGCAGCAAG 122  
QY 183 LeuAspTyrTyrGlnThrMetSerGlyValLysGluAlaAspPheLysAspLeuGluThr 202  
DB 123 CTAGACTATTACCAACCATCTCTGGAGTGAAGAGGCGCGACTTCAAGGATCTTGAGACT 182  
QY 203 GlnProHisMetGlnGluPheArgLeuGlyValArgLeuTyrSerGluGluGlnProGln 222  
DB 183 CAACCCCATATGCAAGATTTCGACTGGAGTGGGACTCTCTACTCAGAGGACACCCACAG 242



DB 662 CAAGAGATACGACAGCAGCAGCTACTGGAAGAAAGAACTGTTTCTCGTTATGATG 721  
 QY 382 lPheGlyleProPheValaspProaspSerTriPThrProGluValleProLysAr 402  
 DB 722 TTTTGAATCCCTTTGTGGATCCGGATTCATGGACTCCAGAAAGAGTATCCCAAGAG 781  
 QY 402 GLeuGlnGluLysGlnLysSerGluAArgGluThrAlaValArgileSerGlnGluile 421  
 DB 782 ATTGCAAGAAACAGAAAGTCAAGACGGGAAACAGCCGTACGCAATCTCCAGGAGATT 839

RESULT 14  
 LOCUS BQ678346 955 bp mRNA linear EST 15-JUL-2002  
 DEFINITION AGENCOURT\_8500887 NIH\_MGC\_112 Homo sapiens cDNA clone IMAGE:6294282  
 5', mRNA sequence.  
 ACCESSION BQ678346  
 VERSION BQ678346.1 GI:21791025  
 KEYWORDS EST.  
 SOURCE Homo sapiens (human)

ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 REFERENCE 1 (Bases 1 to 955)  
 AUTHORS NIH-MGC <http://mgc.nci.nih.gov/>.  
 TITLE National Institutes of Health, Mammalian Gene Collection (MGC)  
 JOURNAL Unpublished (1999)  
 COMMENT Contact: Robert Strausberg, Ph.D.  
 Email: [cgapbs-remail.nih.gov](mailto:cgapbs-remail.nih.gov)  
 Tissue Procurement: DCTD/DTF

cDNA Library Preparation: Rubin Laboratory  
 DNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)  
 DNA Sequencing by: Agencourt Bioscience Corporation  
 Clone distribution: MGC clone distribution information can be  
 found through the I.M.A.G.E. Consortium/LLNL at:  
<http://image.llnl.gov> row: k column: 19  
 Place: LNCN2498  
 High quality sequence stop: 566.  
 Location/Qualifiers  
 1..955  
 /organism="Homo sapiens"  
 /mol\_type="mRNA"  
 /db\_xref="taxon:9606"  
 /clone="IMAGE:6294282"  
 /tissue\_type="melanotic melanoma, cell line"  
 /lab\_host="DH10B (phage-resistant)"  
 /clone\_lib="NIH MGC 112"  
 /note="Organ: skin; Vector: pOTB7; Site 1: XhoI; Site 2:  
 EcoRI; cDNA made by oligo-dT priming. Directionally cloned  
 into EcoRI/XhoI sites using the following 5' adaptor:  
 GCCACGAG(G). Library constructed by Ling Hong in the  
 Laboratory of Gerald M. Rubin (University of California,  
 Berkeley) using ZAP-cDNA synthesis kit (Stratagene) and  
 Superscript II RT (Life Technologies). Note: this is a  
 NIH\_MGC Library."

## FEATURES

source

RESULT 15  
 LOCUS BQ678346 955 bp mRNA linear EST 01-JUN-2000  
 DEFINITION EST376243 MAGE resequences, MAGE Homo sapiens cDNA, mRNA sequence.  
 ACCESSION BQ678346  
 VERSION AW964170.1 GI:8154006  
 KEYWORDS EST.  
 SOURCE Homo sapiens (human)  
 ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 REFERENCE 1 (Bases 1 to 730)  
 AUTHORS Hegde, P., Qi, R., Abernathy, K., Dharap, S., Gaspar, R., Gay, C.,  
 Holt, I. E., Saeed, A. I., Sharov, V., Lee, N. H., Yeatman, T. J. and  
 Quackenbush, J.

## ORIGIN

Alignment Scores:  
 Pred. No.: 1.45e-153 Length: 955  
 Score: 174.00 Matches: 231  
 Percent Similarity: 99.14% Conservative: 0  
 Best Local Similarity: 99.14% Mismatch: 1  
 Query Match: 23.64% Indels: 2  
 DB: 13 Gaps: 0  
 US-10-045-815-4 (1-736) x BQ678346 (1-955)

QY 412 GluThrAlaValArgileSerGlnGluileGlyAsnLeuMetLysGluileGluThrLeu 431  
 DB 1 GAAACAGCGGTACGCATCTCCAGAGATGGGAACCTTATGAAGGAATCGAGACCTT 60  
 QY 432 ValGlnGluLysThrLysGluSerLeuaspValSerArgLeuThrArgGluGlyPro 451

DB 61 GTGGAAGAGAACCAAGGAGTCACTGGATGTGACAGACTGACCCGGGAAGGTGGCCCC 120  
 QY 452 LeuLeuTyrrGluGlylSerLeuThrMetAsnSerLysLeuLeuAsnGlyTyr-GlnAr 471  
 DB 121 CTGCTGTATGAAGGATCAGTCTCCATGAATCAAACTCCTGAATGGTTC-CCAGCG 179  
 QY 471 gValValMetAspGlyValIleSerAspHisGluCysGlnGluLeuGlnArgLeuThrAs 491  
 DB 180 GTGTGTATGAGCGGGTATCTCTGACCAAGAGTGTGAGAGCTGCAGAGACTGACCAA 239  
 QY 491 nValAlaAlaThrSerGlyAspGlyTyrArgGlyGlnThrSerProHisThrProAsnGl 511  
 DB 240 TGTGGCAGCAACCTCAGGAGATGGCTAGCGGGGTGAGACCTCCCCACATACTCCCAATGA 299  
 QY 511 uLysPheTyrrGlyValThrValPheLysAlaLeuLysLeuGlyGlnGluGlyLysValPr 531  
 DB 300 AAAGTTCTATGGTGTACCTCTTCARAGCCCTCAAGCTGGGGCAAGGAAGCAAGTTCC 359  
 QY 531 oLeuGlnSerAlaHisLeuTyrrAsnValThrGluLysValArgArgileMetGluSe 551  
 DB 360 TCTGCGAGAGTGCACCTGTACTCAACGTCAGCGAAGAGTGGCGGCATCATGGAGTC 419  
 QY 551 rTyrrPheArgLeuAspThrProLeuTyrrPheSerTyrrSerHisLeuValCysArgThrAl 571  
 DB 420 CTACTTCGCGCTGGATGACGCCCTCTACTTTTCTTACTCTCTATCTGGTGGCCGACTGC 479  
 QY 571 aileGluGluValGlnAlaGluAArgLysAspSerHisProValHisValAspAsnCy 591  
 DB 480 CATCGAAGAGTCCAGGCAGAGAGAGATAGTAGTCATCCAGTCCAGTGGACAACTG 539  
 QY 591 sileLeuAsnAlaGluThrLeuValCysValLysGluPr-oProAlaTyrrThrPheArgAs 611  
 DB 540 CATCTGTAATGCGAGACCTCTGTGTGTCAAGAGAGCCCGCAGCTACACCTTCCCGGA 599  
 QY 611 pTyrrSerAlaIleLeuTyrrLeuAsnGlyAspPheAspGlyGlyAsnPheTyrrPheThrGl 631  
 DB 600 CTACAGCGCCATCCCTTTTACCTAAATGGGAGCTTCGATGGCGGAACCTTTTATTCACGTA 659  
 QY 631 uLeuAspAlaLysThrValThrAlaGluValGlnPro 643  
 DB 660 ACTGATGCCAAGACCGTGACGGCAGAGAGTGCAGCCT 696

RESULT 15  
 LOCUS AW964170 730 bp mRNA linear EST 01-JUN-2000  
 DEFINITION EST376243 MAGE resequences, MAGE Homo sapiens cDNA, mRNA sequence.  
 ACCESSION AW964170  
 VERSION AW964170.1 GI:8154006  
 KEYWORDS EST.  
 SOURCE Homo sapiens (human)  
 ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 REFERENCE 1 (Bases 1 to 730)  
 AUTHORS Hegde, P., Qi, R., Abernathy, K., Dharap, S., Gaspar, R., Gay, C.,  
 Holt, I. E., Saeed, A. I., Sharov, V., Lee, N. H., Yeatman, T. J. and  
 Quackenbush, J.

Assessment of gene expression patterns in a model of colon tumor  
 metastasis using a 19,200 element cDNA microarray  
 Unpublished (2000)

CONTACT: John Quackenbush  
 The Institute for Genomic Research  
 9712 Medical Center Dr., Rockville, MD 20850, USA  
 Tel: 301 838 3528  
 Fax: 301 838 0208  
 Email: [johnq@tigr.org](mailto:johnq@tigr.org)  
 Plate: 198  
 Seq primer: Reverse.  
 Location/Qualifiers  
 1..730  
 /organism="Homo sapiens"  
 /mol\_type="mRNA"  
 /db\_xref="taxon:9606"

## FEATURES

source

/clone\_lib="WAGE resequences, MAGH"  
/note="Vector: pBluescriptSKm"

ORIGIN

Alignment Scores: 5.11e-148 Length: 730  
Pred. No.: 168.00 Matches: 168  
Score: 100.00% Conservative: 0  
Percent Similarity: 100.00% Mismatches: 0  
Best Local Similarity: 100.00% Indels: 0  
Query Match: 22.83% Gaps: 0  
DB: 10

US-10-045-815-4 (1-736) x AW964170 (1-730)

Qy 569 ArgThrAlaIleGluGluValGlnAlaGluArgLysAspSerHisProValHisVal 588  
Db 1 CGCACTGCCATCGAGAGGTCCAGGCGAGAGAGAGATGATGATCATCCAGTCCACGTG 60  
Qy 589 AspAsnCysIleLeuAsnAlaGluThrLeuValCysValLysGluProAlaIleThr 608  
Db 61 GACAACTGCATCTCGATGCCGAGACCCCTGCTGTCTCAAGAGAGCCCCCAGGCTACAC 120  
Qy 609 PheArgAspTyrSerAlaIleLeuTyrLeuAsnGlyAspPheAspGlyGlyAsnPheTyr 628  
Db 121 TTCGCGACCTACAGCGCCATCTTTACCTAAATGGGGACTTCGATGGCGGAAACTTTTAT 180  
Qy 629 PheThrGluLeuAspAlaLysThrValThrAlaGluValGlnProGlnCysGlyArgAla 648  
Db 181 TTCACCTGAACCTGGATGCCAAGACCGTGCAGGCGAGAGGTGCAGCTCAGTGTGGAAGAGCC 240  
Qy 649 ValGlyPheSerSerGlyThrGluAsnProHisGlyValLysAlaValThrArgGlyGln 668  
Db 241 GTGGGATTCCTTCAGGCACTGAAACCCACATGGAGTGAAGGCTGTACCCAGGGGGCAG 300  
Qy 669 ArgCysAlaIleAlaLeuTyrPheThrLeuAspProArgHisSerGluArgAspArgVal 688  
Db 301 CGCTGTGCCATCGCCCTGTGGTTACCCCTGGACCCCTGCACACAGCGAGCGGGACAGGGTG 360  
Qy 689 GlnAlaAspAspLeuValLysMetLeuPheSerProGluGluMetAspLeuSerGlnGlu 708  
Db 361 CAGGCAGATGACCTGTGTGAAGATGCTCTTCAGCCCGAGAGAGATGGACCTCTCCAGGAG 420  
Qy 709 GlnProLeuAspAlaGlnGlnGlyProGluProAlaGlnGluSerLeuSerGlySer 728  
Db 421 CAGCCCTGTGATGCCCGAGCGGGCCCCCGAACCCTGCACAGAGATCTCTCTCAGGCAGT 480  
Qy 729 GluSerLysProLysAspGluLeu 736  
Db 481 GATCGAAGCCCAAGGATGAGCTA 504

Search completed: July 18, 2004, 23:26:20  
Job time : 6563 secs

This Page Blank (uspto)



GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: July 18, 2004, 07:57:32 ; Search time 105 Seconds  
(without alignments)  
1980.523 Million cell updates/sec

Title: US-10-045-815-4

Perfect score: 3870

Sequence: 1 MAVRALKLTLLAVVAAS.....PPEPAQESLSGSESKPKDEL 736

Scoring table:

Gapop 10.0 , Gapext 0.5

Searched: 1586107 seqs, 282547505 residues

Total number of hits satisfying chosen parameters: 1586107

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : A\_Geneseq\_29Jan04.\*

1: Geneseqp1980s.\*

2: Geneseqp1990s.\*

3: Geneseqp2000s.\*

4: Geneseqp2001s.\*

5: Geneseqp2002s.\*

6: Geneseqp2003as.\*

7: Geneseqp2003bs.\*

8: Geneseqp2004s.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match %	Length	DB ID	Description
1	3870	100.0	736	3 AAB36392	Aab36392 Human tum
2	3861	99.8	736	5 ABP69129	Abp69129 Human pol
3	3846	99.4	736	4 AAB93215	Aab93215 Human pro
4	3846	99.4	736	4 AAB88373	Aab88373 Human mem
5	3846	99.4	736	5 AAU09861	Aau09861 Novel hum
6	3844	99.3	736	4 AAB93142	Aab93142 Human pro
7	3801	98.2	806	5 ABG66709	Abg66709 Human nov
8	3801	98.2	806	5 ABG66686	Abg66686 Human nov
9	3645	94.2	706	5 ABP69128	Abp69128 Human pol
10	3582	92.6	804	4 AAB94482	Aab94482 Human pro
11	3572	92.3	708	5 AAU09860	Aau09860 Novel hum
12	3316	85.7	747	3 AAB36393	Aab36393 Mouse tum
13	3316	85.7	747	7 ABR63160	Abr63160 Murine po
14	3115	80.5	592	4 AAE04244	Aae04244 Human gen
15	2384	61.6	542	3 AAB36394	Aab36394 Mouse tum
16	2147	55.5	408	5 ABP69127	Abp69127 Human pol
17	2065	53.4	400	5 ABG73582	Abg73582 Human leu
18	1911	49.4	363	3 AAB36391	Aab36391 Human tum
19	1800	46.5	359	4 AAE04200	Aae04200 Human gen
20	1800	46.5	359	5 ABG64502	Abg64502 Human alb
21	1751	45.2	708	6 ADA54992	Ada54992 Human pro
22	1747	45.1	708	7 ABR63159	Abr63159 Human pro
23	1722	44.5	337	4 AAE04246	Aae04246 Human gen
24	1473	38.1	736	6 ABU52322	Abu52322 Human GPC
25	1409.5	36.4	527	4 AAB93094	Aab93094 Human pro

26	1179.5	30.5	551	6 AAO30313	Aao30313 Human pro
27	1123	25.0	217	3 AAB42734	Ab42734 Human ORF
28	977	25.2	182	6 ABU70516	Abu70516 Human adi
29	906.5	23.4	282	7 ADB82712	Ad82712 Human pro
30	836.5	21.6	267	4 AAU17171	Aau17171 Novel sig
31	836.5	21.6	267	4 AAU87588	Aau87588 Novel cen
32	836.5	21.6	267	7 ADB93879	Ad93879 Human nov
33	790	20.4	227	4 ABG16396	Abg16396 Novel hum
34	657	17.0	173	4 AAE04217	Aae04217 Human gen
35	657	17.0	173	5 ABG64501	Abg64501 Human alb
36	612	15.8	219	4 AAU87297	Aau87297 Novel cen
37	572	14.9	401	4 AAB93152	Aab93152 Human pro
38	572	14.8	401	7 ADE60315	Ad60315 Human pro
39	572	14.8	401	7 ADE60327	Ad60327 Human pro
40	572	14.8	401	7 ADE60319	Ad60319 Human pro
41	572	14.8	401	7 ADE60323	Ad60323 Human pro
42	572	14.8	401	7 ADE59277	Ad59277 Human pro
43	567	14.7	401	2 AAU17219	Aay17219 Human sec
44	567	14.7	401	4 AAU38989	Aau38989 Human sec
45	567	14.7	401	5 ABB55698	Abb55698 Human pol

## ALIGNMENTS

### RESULT 1

AAB36392

ID AAB36392 standard; protein; 736 AA.

XX

AC AAB36392;

XX

DT 27-FEB-2001 (first entry)

XX

DE Human tumour suppressor Gros1-S protein SEQ ID NO:4.

XX

KW Tumour suppressor; Gros1-L; Gros1-S; cell proliferation; regulation;

KW cancer; cytostatic; gene therapy.

XX

OS Homo sapiens.

XX

PN WO200065047-A1.

XX

PD 02-NOV-2000.

XX

PF 26-APR-2000; 2000WO-JF002731.

XX

PR 26-APR-1999; 99JP-00118806.

XX

PA (CHUG-) CHUGAI RES INST MOLECULAR MEDICINE INC.

XX

PI Wadhwa R, Sugihara T, Yoshida A;

XX

DR WPI; 2000-687340/67.

XX

DR N-PSDB; AAC64725.

XX

PT Full-length tumour suppressor genes encoding Gros1-L, Gros1-S and mouse homologs participating in regulation of cell proliferation, useful in development of preventives and remedies of cancer.

XX

PS Claim 1; Page 78-83; 114pp; Japanese.

XX

CC The present sequence represents the human tumour suppressor designated Gros1-S. Gros1-L and Gros1-S have cytostatic activity and can be used in gene therapy. Gros1-L and Gros1-S genes are useful in the development of drugs used to treat and prevent cancer

XX

SQ Sequence 736 AA;

Query Match 100.0%; Score 3870; DB 3; Length 736;

Best Local Similarity 100.0%; Pred.No. 0;

Matches 736; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

1 MAVRALKLTLLAVVAASQAQVESEAGWGMVTPDLLFAEGTAAAYARGDWPVVLSMER 60

Db 1 MAVRAKLLTLLAVVAAASQAQVESEAGWGVTDDLPAEGTAAYARGDWPVGLSMER 60  
 Qy 61 ALRSRAALRALRLRCRTCAADFFWELDPDMSPPAQASGAGALRDLSPFGGLLRRACL 120  
 Db 61 ALRSRAALRALRLRCRTCAADFFWELDPDMSPPAQASGAGALRDLSPFGGLLRRACL 120  
 Qy 121 RRCLGPPAAHSLSEMELEFRKRSPYNYLQVAYFKINKLEKAVAAAHTEFFVGNPEHMEMQ 180  
 Db 121 RRCLGPPAAHSLSEMELEFRKRSPYNYLQVAYFKINKLEKAVAAAHTEFFVGNPEHMEMQ 180  
 Qy 181 QNLDDYQTMGSKVKEADFKDLETPHMQEFLGRLVLYSEEQPQAVPHLEALQYFVAYE 240  
 Db 181 QNLDDYQTMGSKVKEADFKDLETPHMQEFLGRLVLYSEEQPQAVPHLEALQYFVAYE 240  
 Qy 241 ECRALCEGPDYDGVNLYEYNADLFOAITDHYIQLVNLCKQNCVTELSHPSREKPFDFL 300  
 Db 241 ECRALCEGPDYDGVNLYEYNADLFOAITDHYIQLVNLCKQNCVTELSHPSREKPFDFL 300  
 Qy 301 PSHYNYLOFAYYINIGNYTQAGECAKTYLLFFPNDVNMQLAYYAAAMLGEBHTRSIGPRE 360  
 Db 301 PSHYNYLOFAYYINIGNYTQAGECAKTYLLFFPNDVNMQLAYYAAAMLGEBHTRSIGPRE 360  
 Qy 361 SAKEYRQSRSLLEKELLFFAYDVGIPFVDPDSWTPPEVPIKRLQEKOKSERETAVRISQE 420  
 Db 361 SAKEYRQSRSLLEKELLFFAYDVGIPFVDPDSWTPPEVPIKRLQEKOKSERETAVRISQE 420  
 Qy 421 IGNLMKETETLVEBKTESLVSRLTREGGPLLVEGLSLTWSKLLNGYQVWMDGVISD 480  
 Db 421 IGNLMKETETLVEBKTESLVSRLTREGGPLLVEGLSLTWSKLLNGYQVWMDGVISD 480  
 Qy 481 HECBELQRLTNAATSGDGYRGQTSPTPNKFKYGVTVFVKALKLGQSGKVPLOSAHLYYN 540  
 Db 481 HECBELQRLTNAATSGDGYRGQTSPTPNKFKYGVTVFVKALKLGQSGKVPLOSAHLYYN 540  
 Qy 541 VTEKVRIMESYFRDLTPLYFSYSHLVCRTAIEVQAEKDDSHPHVHVCNLIINAETLVC 600  
 Db 541 VTEKVRIMESYFRDLTPLYFSYSHLVCRTAIEVQAEKDDSHPHVHVCNLIINAETLVC 600  
 Qy 601 VKEPPAYTFRYSAILYINGDFDGNFYFTLDAKTYTAEVQPCGQRAVGFSSGTENPHG 660  
 Db 601 VKEPPAYTFRYSAILYINGDFDGNFYFTLDAKTYTAEVQPCGQRAVGFSSGTENPHG 660  
 Qy 661 VKATRGORCAIALWFTLDPRHSERDRVQADDLVKMLFSPREMDLSOEPLDAQQGPPEP 720  
 Db 661 VKATRGORCAIALWFTLDPRHSERDRVQADDLVKMLFSPREMDLSOEPLDAQQGPPEP 720  
 Qy 721 AQESLSGSESXPKDEL 736  
 Db 721 AQESLSGSESXPKDEL 736

## RESULT 2

ABP69129

ID ABP69129 standard; protein; 736 AA.

XX AC

XX AC

XX ABP69129;

DT 20-JAN-2003 (first entry)

XX Human polypeptide SEQ ID NO 1176.

DE DE

KW Human; genome mapping; gene therapy; food supplement; virus; fungus;  
 cell-proliferative disorder; neurodegenerative disease; bacterial;  
 Parkinson's disease; Alzheimer's disease; autoimmune disease;  
 multiple sclerosis; diabetes; genetic disorder; wound; burn; infection;  
 arthritis; cytostatic; immunomodulator; nootropic; neuroprotective;  
 antiparkinsonian; antidiabetic; immunosuppressive; dermatological;  
 haemostatic; vulnery; fungicide; antibacterial; virucide; protozoacide;  
 antiarthritic.

XX OS

XX Homo sapiens.

XX

PN WO200270539-A2.  
 PD 12-SEP-2002.  
 PF 05-MAR-2002; 2002WO-US005095.  
 XX 05-MAR-2001; 2001US-00799451.  
 XX (HYSE-) HYSEQ INC.  
 PA Tang YT, Zhou P, Goodrich RW, Asundi V, Zhang J, Zhao QA, Ren F;  
 PI Xue AJ, Yang Y, Ma Y, Yamazaki V, Chen R, Wang Z, Ghosh M;  
 PI Wehrman T, Wang J, Wang D, Drmanac RT;  
 XX WPI: 2002-759812/82.  
 DR N-PSDB; AB211346.  
 XX New polynucleotides comprising sequences assembled from expressed  
 PT sequence tags (ESTs), useful for treating cell-proliferative,  
 PT neurodegenerative, autoimmune, genetic, myeloid or lymphoid, or platelet  
 PT or coagulation disorders.  
 XX Claim 9; SEQ ID NO 1176; 1012pp + Sequence Listing; English.  
 PS The invention relates to an isolated polynucleotide (I) comprising a  
 CC nucleotide sequence selected from any of 948 sequences (AB21119-  
 CC AB212066) or their mature protein coding portion, active domain coding  
 CC protein or complementary sequences. The polynucleotides are useful for  
 CC identifying expressed genes or for physical mapping of human genome. The  
 CC encoded polypeptides (ASP68902-ASP69849) are useful as molecular weight  
 CC markers, as a food supplement, for generating antibodies, in medical  
 CC imaging, screening and diagnostic assays and for treating cell-  
 CC proliferative disorders (cancer), neurodegenerative diseases (Parkinson's  
 CC or Alzheimer's disease), autoimmune diseases (multiple sclerosis,  
 CC diabetes, lupus) genetic disorders, myeloid or lymphoid disorders,  
 CC platelet or coagulation disorders, wound, burns, incision, ulcers, liver  
 CC or lung fibrosis infections (bacterial, viral, fungal, parasitic),  
 CC arthritis, etc. Note: The sequence data for this patent did not form part  
 CC of the printed specification, but was obtained in electronic format  
 CC directly from WIFO at ftp.wipo.int/pub/published\_pct\_sequences  
 XX Sequence 736 AA;

Query Match 99.8%; Score 3861; DB 5; Length 736;

Best Local Similarity 99.7%; Pred: No. 0;

Matches 734; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Qy 1 MAVRAKLLTLLAVVAAASQAQVESEAGWGVTDDLPAEGTAAYARGDWPVGLSMER 60  
 Db 1 MAVRAKLLTLLAVVAAASQAQVESEAGWGVTDDLPAEGTAAYARGDWPVGLSMER 60  
 Qy 61 ALRSRAALRALRLRCRTCAADFFWELDPDMSPPAQASGAGALRDLSPFGGLLRRACL 120  
 Db 61 ALRSRAALRALRLRCRTCAADFFWELDPDMSPPAQASGAGALRDLSPFGGLLRRACL 120  
 Qy 121 RRCLGPPAAHSLSEMELEFRKRSPYNYLQVAYFKINKLEKAVAAAHTEFFVGNPEHMEMQ 180  
 Db 121 RRCLGPPAAHSLSEMELEFRKRSPYNYLQVAYFKINKLEKAVAAAHTEFFVGNPEHMEMQ 180  
 Qy 181 QNLDDYQTMGSKVKEADFKDLETPHMQEFLGRLVLYSEEQPQAVPHLEALQYFVAYE 240  
 Db 181 QNLDDYQTMGSKVKEADFKDLETPHMQEFLGRLVLYSEEQPQAVPHLEALQYFVAYE 240  
 Qy 241 ECRALCEGPDYDGVNLYEYNADLFOAITDHYIQLVNLCKQNCVTELSHPSREKPFDFL 300  
 Db 241 ECRALCEGPDYDGVNLYEYNADLFOAITDHYIQLVNLCKQNCVTELSHPSREKPFDFL 300  
 Qy 301 PSHYNYLOFAYYINIGNYTQAGECAKTYLLFFPNDVNMQLAYYAAAMLGEBHTRSIGPRE 360  
 Db 301 PSHYNYLOFAYYINIGNYTQAGECAKTYLLFFPNDVNMQLAYYAAAMLGEBHTRSIGPRE 360  
 Qy 361 SAKEYRQSRSLLEKELLFFAYDVGIPFVDPDSWTPPEVPIKRLQEKOKSERETAVRISQE 420

Db 361 SAKEYRQSLLEKELFFAYDVDFGIFVDDSWTPPEVTPKRLQEKQKSERETAVRISQE 420  
Qy 421 IGNLMKEIETLVBEKTKESLDVSLRTREGGPLYEGISLTMSKLLNGYQRVVMDGVISD 480  
Db 421 IGNLMKEIETLVBEKTKESLDVSLRTREGGPLYEGISLTMSKLLNGYQRVVMDGVISD 480  
Qy 481 HECQELQRLTNVAATSGDGYRGQTSPTHTNEKEFGYTVFKALKLGGEGKVPLOSALHLYN 540  
Db 481 HECQELQRLTNVAATSGDGYRGQTSPTHTNEKEFGYTVFKALKLGGEGKVPLOSALHLYN 540  
Qy 541 VTEKVRIMESYFRLDTPLFYFSYSHLVCRTAIEVQAEKRDSDHPVHVNDNCILNAETLVC 600  
Db 541 VTEKVRIMESYFRLDTPLFYFSYSHLVCRTAIEVQAEKRDSDHPVHVNDNCILNAETLVC 600  
Qy 601 VKPPPAYTFRDYSAILYLNGDFDGGNFYFTLDAKTVAEVQPCQRAVGFSSGTENPHG 660  
Db 601 VKPPPAYTFRDYSAILYLNGDFDGGNFYFTLDAKTVAEVQPCQRAVGFSSGTENPHG 660  
Qy 661 VKAVTGRGRCALWFTLDPHRSERDRVQADDLVKMLFSPPEMDLSQEQPLDAQQQPPRP 720  
Db 661 VKAVTGRGRCALWFTLDPHRSERDRVQADDLVKMLFSPPEMDLSQEQPLDAQQQPPRP 720  
Qy 721 AQESLSGSESKPKDEL 736  
Db 721 AQESLSGSESKPKDEL 736

RESULT 3  
AAB93215  
ID AAB93215 standard; protein; 736 AA.  
AC AAB93215;  
XX 26-JUN-2001 (first entry)  
DE Human protein sequence SEQ ID NO:12194.  
XX Human; primer; detection; diagnosis; antisense therapy; gene therapy.  
XX Homo sapiens.  
XX EPI074617-A2.  
XX 07-FEB-2001.  
XX 28-JUL-2000; 2000EP-00116126.  
XX 29-JUL-1999; 99JP-00248036.  
XX 27-AUG-1999; 99JP-00300253.  
XX 11-JAN-2000; 2000JP-00118776.  
XX 02-MAY-2000; 2000JP-00183767.  
XX 09-JUN-2000; 2000JP-00241899.  
XX (HELI-) HELIX RES INST.  
XX Ota T, Isogai T, Nishikawa T, Hayashi K, Saito K, Yamamoto J;  
PI Ishii S, Sugiyama T, Wakamatsu A, Nagai K, Otsuki T;  
XX WPI; 2001-318749/34.  
XX Primer sets for synthesizing polynucleotides, particularly the 5602 full-length cDNAs defined in the specification, and for the detection and/or diagnosis of the abnormality of the proteins encoded by the full-length cDNAs.  
XX Claim 8; SEQ ID NO 12194; 2537pp + Sequence Listing; English.  
XX The present invention describes primer sets for synthesizing 5602 full-length cDNAs defined in the specification. Where a primer set comprises: (a) an oligo-dT primer and an oligonucleotide complementary to the complementary strand of a polynucleotide which comprises one of the 5602 nucleotide sequences defined in the specification, where the oligonucleotide comprises at least 15 nucleotides; or (b) a combination

of an oligonucleotide comprising a sequence complementary to the complementary strand of a polynucleotide which comprises a 5'-end sequence and an oligonucleotide comprising a sequence complementary to a polynucleotide which comprises a 3'-end sequence, where the oligonucleotide comprises at least 15 nucleotides and the combination of the 5'-end sequence/3'-end sequence is selected from those defined in the specification. The primer sets can be used in antisense therapy and in gene therapy. The primers are useful for synthesizing polynucleotides, particularly full-length cDNAs. The primers are also useful for the detection and/or diagnosis of the abnormality of the proteins encoded by the full-length cDNAs. The primers allow obtaining of the full-length cDNAs easily without any specialised methods. AAH03166 to AAH13628 and AAH13633 to AAH19742 represent human cDNA sequences; AAB92446 to AAB95893 represent human amino acid sequences; and AAH13629 to AAH13632 represent oligonucleotides, all of which are used in the exemplification of the present invention  
XX  
SQ Sequence 736 AA;  
Query Match 99.4%; Score 3846; DB 4; Length 736;  
Best Local Similarity 99.6%; Pred. No. 0;  
Matches 733; Conservative 0; Mismatches 3; Indels 0; Gaps 0;  
Qy 1 MAVRALKLLTLLAVVAAASQAQVESEAGWGMVTPDLLFAEGTAAYARGDPPGVLSMER 60  
Db 1 MAVRALKLLTLLAVVAAASQAQVESEAGWGMVTPDLLFAEGTAAYARGDPPGVLSMER 60  
Qy 61 ALRSPALRALRLCRTOCAADFPWELDPDWSPPAQASGAGALADLSFFGGLLRAACL 120  
Db 61 ALRSPALRALRLCRTOCAADFPWELDPDWSPPAQASGAGALADLSFFGGLLRAACL 120  
Qy 121 RRLGPPAAHSLSEMELEFRKSPYNYLOVAYFKINKLEKAVAAAHFFVGNPEHMEMQ 180  
Db 121 RRLGPPAAHSLSEMELEFRKSPYNYLOVAYFKINKLEKAVAAAHFFVGNPEHMEMQ 180  
Qy 181 QNLDYQTMGSKVKEADFXDLETPHMQBFRGLVRLYSEEQPQEAHPLEALQYFVAYE 240  
Db 181 QNLDYQTMGSKVKEADFXDLETPHMQBFRGLVRLYSEEQPQEAHPLEALQYFVAYE 240  
Qy 241 ECRALCEGPYDGYNYLEYNADLFOAITDHYIQLVNCQKNCVTELASHPSREKPFDFL 300  
Db 241 ECRALCEGPYDGYNYLEYNADLFOAITDHYIQLVNCQKNCVTELASHPSREKPFDFL 300  
Qy 301 PSHYNYLOFAYYNYIGNYTQAGECAKTYLLFFPNDEVMQNLAYYAAMLGEEHTRSIGPRE 360  
Db 301 PSHYNYLOFAYYNYIGNYTQAGECAKTYLLFFPNDEVMQNLAYYAAMLGEEHTRSIGPRE 360  
Qy 361 SAKEYRQSLLEKELFFAYDVDFGIFVDDSWTPPEVTPKRLQEKQKSERETAVRISQE 420  
Db 361 SAKEYRQSLLEKELFFAYDVDFGIFVDDSWTPPEVTPKRLQEKQKSERETAVRISQE 420  
Qy 421 IGNLMKEIETLVBEKTKESLDVSLRTREGGPLYEGISLTMSKLLNGYQRVVMDGVISD 480  
Db 421 IGNLMKEIETLVBEKTKESLDVSLRTREGGPLYEGISLTMSKLLNGYQRVVMDGVISD 480  
Qy 481 HECQELQRLTNVAATSGDGYRGQTSPTHTNEKEFGYTVFKALKLGGEGKVPLOSALHLYN 540  
Db 481 HECQELQRLTNVAATSGDGYRGQTSPTHTNEKEFGYTVFKALKLGGEGKVPLOSALHLYN 540  
Qy 541 VTEKVRIMESYFRLDTPLFYFSYSHLVCRTAIEVQAEKRDSDHPVHVNDNCILNAETLVC 600  
Db 541 VTEKVRIMESYFRLDTPLFYFSYSHLVCRTAIEVQAEKRDSDHPVHVNDNCILNAETLVC 600  
Qy 601 VKPPPAYTFRDYSAILYLNGDFDGGNFYFTLDAKTVAEVQPCQRAVGFSSGTENPHG 660  
Db 601 VKPPPAYTFRDYSAILYLNGDFDGGNFYFTLDAKTVAEVQPCQRAVGFSSGTENPHG 660  
Qy 661 VKAVTGRGRCALWFTLDPHRSERDRVQADDLVKMLFSPPEMDLSQEQPLDAQQQPPRP 720  
Db 661 VKAVTGRGRCALWFTLDPHRSERDRVQADDLVKMLFSPPEMDLSQEQPLDAQQQPPRP 720  
Qy 721 AQESLSGSESKPKDEL 736  
Db 721 AQESLSGSESKPKDEL 736

Db 721 AQESLSGSESKPKDEL 736

RESULT 4

AA883373

ID AAB88373 standard; protein; 736 AA.

XX AC AAB88373;

XX DT 23-MAY-2001 (first entry)

XX DE Human membrane or secretory protein clone PSEC0109.

XX KW Human; secretory protein; membrane protein; vaccine; gene therapy;

XX KW rheumatoid arthritis; diabetes.

XX OS Homo sapiens.

XX FN EPI067182-A2.

XX PD 10-JAN-2001.

XX PF 07-JUL-2000; 2000EP-00114090.

XX PR 08-JUL-1999; 99JP-00194179.

XX PR 11-JAN-2000; 2000JP-00118775.

XX PR 02-MAY-2000; 2000JP-00183766.

XX PA (HELI-) HELIX RES INST.

XX PI Ota T, Isogai T, Nishikawa T, Kawai Y, Sugiyama T, Hayashi K;

XX WPI; 2001-093989/11.

XX DR N-PSDB; AAF93800.

XX PT Nucleic acids encoding secretory proteins/membrane proteins, useful in

XX PT gene therapy or as candidate target molecules in drug development.

XX PS Claim 1; SEQ ID NO 114; 609pp + Sequence Listing; English.

XX CC This invention relates to nucleic acid sequences AAF93744 - AAF93916

XX CC which encode human secretory or membrane proteins represented by AAB88317

XX CC - AAB88419. Included in the invention are primers AAF93917 - AAF94295 and

XX CC AAF62232 - AAF62235 which are used to isolate the cDNA sequences of the

XX CC invention. The invention also includes methods for the production of

XX CC antibodies directed against the proteins, and cDNA sequences, which can

XX CC be used in vaccines. The polynucleotide sequences can be used in gene

XX CC therapy. The polynucleotide sequences and the proteins they encode may be

XX CC used in the prevention, treatment and diagnosis of diseases associated

XX CC with inappropriate secretory protein/membrane protein expression. The

XX CC nucleic acids and complementary sequences may also be used as DNA probes

XX CC in diagnostic assays (e.g. polymerase chain reactions (PCR)) to detect

XX CC and quantitate the presence of similar nucleic acid sequences in samples.

XX CC They may also be used to study the expression and function of secretory

XX CC proteins/membrane polypeptides and their role in metabolism. The

XX CC polypeptides may be used as antigens in the production of antibodies

XX CC against them and in assays to identify modulators (agonists and

XX CC antagonists) of expression and activity. The antibodies and antagonists

XX CC may also be used as therapeutic agents to down regulate expression and

XX CC activity. The antibodies may also be used as diagnostic agents for

XX CC detecting the presence of the polypeptides in samples (e.g. by enzyme

XX CC linked immunosorbent assay (ELISA). Examples of diseases which may be

XX CC treated include rheumatoid arthritis and diabetes

XX SQ Sequence 736 AA;

Query Match 99.4%; Score 3846; DB 4; Length 736;

Best Local Similarity 99.6%; Pred. No. 0;

Matches 733; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 1 MAVRALKLLTLLAVVAASQAQVESEAGMGWVTPDLLFAEGTAAYARGDWPQVLSMER 60

DB 1 MAVRALKLLTLLAVVAASQAQVESEAGMGWVTPDLLFAEGTAAYARGDWPQVLSMER 60

QY 61 ALRSRAALRALRLRCRTCCAADFPWELDPDWSPPAQASGAGALRDLSFFGGLLRRACL 120

DB 61 ALRSRAALRALRLRCRTCCAADFPWELDPDWSPPAQASGAGALRDLSFFGGLLRRACL 120

QY 121 RCLGPPAAHSLSEEMELEFRKRSYNYLVAYFKINKLEKAVAAHTFFVGNBEHMEMQ 180

DB 121 RCLGPPAAHSLSEEMELEFRKRSYNYLVAYFKINKLEKAVAAHTFFVGNBEHMEMQ 180

QY 181 QNLDDYQTMISGVKCEADFKDLETPHMQEFLRGVRLYSEEQPCEAVPHLEALQYFFVAYE 240

DB 181 QNLDDYQTMISGVKCEADFKDLETPHMQEFLRGVRLYSEEQPCEAVPHLEALQYFFVAYE 240

QY 241 ECRALCEGPDYDGYNYLYNADLFOAITDHYIQVLNCKQNCVTELASHSPREKPFEDFL 300

DB 241 ECRALCEGPDYDGYNYLYNADLFOAITDHYIQVLNCKQNCVTELASHSPREKPFEDFL 300

QY 301 PSHYNYLOFAYNYIGNYTQAGECAKTYLLFPNDEVNMNLAYYAAMLGEEHTSIGPRE 360

DB 301 PSHYNYLOFAYNYIGNYTQAGECAKTYLLFPNDEVNMNLAYYAAMLGEEHTSIGPRE 360

QY 361 SAKYRQSRSLLEKELLFFAYDVFGIPFVDPDSWTPPEEVI PKRLOEKQKSERETAVRISQE 420

DB 361 SAKYRQSRSLLEKELLFFAYDVFGIPFVDPDSWTPPEEVI PKRLOEKQKSERETAVRISQE 420

QY 421 IGNLMKEITLVEEKTESLDVSRLTREGGPIILYEGISLTNSKLLNGYQRVMDGVISD 480

DB 421 IGNLMKEITLVEEKTESLDVSRLTREGGPIILYEGISLTNSKLLNGYQRVMDGVISD 480

QY 481 HECQLQLRLTNVAATSGDYGQTSPTPNKPFYGVTVFKALKLCQEGKVPLOSAHLYYN 540

DB 481 HECQLQLRLTNVAATSGDYGQTSPTPNKPFYGVTVFKALKLCQEGKVPLOSAHLYYN 540

QY 541 VTEKYRRIMESYFRDLTPLYFSYSHLVCRVTAIEVQAEKDDSHPVHVDNCLNAETLVC 600

DB 541 VTEKYRRIMESYFRDLTPLYFSYSHLVCRVTAIEVQAEKDDSHPVHVDNCLNAETLVC 600

QY 601 VKEPPAYTFRDYSATLYLNGDFDGNFVETELDAKTVTAEVQPCGAVGSSGTENPHG 660

DB 601 VKEPPAYTFRDYSATLYLNGDFDGNFVETELDAKTVTAEVQPCGAVGSSGTENPHG 660

QY 661 VKAVTRGQRCALWFTLDPHRSERDRVQADDLVKMLFSPPEMDLSQEPDLDAQGPPPEP 720

DB 661 VKAVTRGQRCALWFTLDPHRSERDRVQADDLVKMLFSPPEMDLSQEPDLDAQGPPPEP 720

QY 721 AQESLSGSESKPKDEL 736

DB 721 AQESLSGSESKPKDEL 736

RESULT 5

AAU09861

ID AAU09861 standard; protein; 736 AA.

XX AC AAU09861;

XX DT 26-FEB-2002 (first entry)

XX DE Novel human secreted protein #2.

XX KW Secreted protein; cytostatic; immunosuppressive; vulnary; vaccine;

XX KW antinflammatory; neuroprotective; nephrotropic; cardiovascular; human;

XX KW cancer; autoimmune disease; wound healing disorder; infection;

XX KW haematopoietic disorder; inflammatory disorder; infertility;

XX KW neurological disease; psychiatric disease; cardiovascular disease;

XX KW respiratory disease; renal; gastrointestinal.

XX OS Homo sapiens.

XX FN WO200179454-A1.

XX PD 25-OCT-2001.

PF 11-APR-2001; 2001WO-US011797.  
XX 13-APR-2000; 2000US-0196603P.  
PR 24-APR-2000; 2000US-0199417P.  
XX (SMIK ) SMITHKLINE BEECHAM CORP.  
PA (SMIK ) SMITHKLINE BEECHAM PLC.  
XX Agarwal P, Murdoch PR, Rizvi SK, Smith RF, Xiang Z;  
PI WPI; 2002-061975/08.  
DR N-PSDB; AAS17573.  
XX  
PT New secreted proteins or polypeptides, useful for treating e.g. cancer,  
PT autoimmune diseases, wound healing disorder, infections, hematopoietic  
PT disorders, inflammatory disorders, infertility, cancer.  
XX  
PS Claim 1; Page 54-56; 92pp; English.  
XX  
CC The invention relates to an isolated novel secreted polypeptide (I) and  
CC polynucleotide (II). (I) and (II) are useful for treating cancer,  
CC autoimmune diseases, wound healing disorder, infections, hematopoietic  
CC disorders, inflammatory disorders, infertility, neurological and  
CC psychiatric diseases, cardiovascular diseases, respiratory diseases,  
CC renal diseases, or gastrointestinal diseases. These may also be used to  
CC treat diseases, abnormalities and disorders caused by abnormal  
CC expression, production, function and/or metabolism of the genes, as  
CC vaccines for inducing immunological response in a mammal, and in  
CC screening methods for detecting the effect of added compounds on the  
CC production of mRNA and polypeptide in cells. The polypeptides can be used  
CC as immunogens to produce antibodies immunospecific for the polypeptides,  
CC and to identify membrane-bound or soluble receptors. The polynucleotides  
CC may be used as diagnostic reagents, in chromosome localisation studies,  
CC and in tissue expression studies. The present sequence represents the  
CC amino acid sequence of novel human secreted protein #2  
XX  
SQ Sequence 736 AA;  
Query Match 99.4%; Score 3846; DB 5; Length 736;  
Best Local Similarity 99.6%; Pred. No. 0;  
Matches 733; Conservative 0; Mismatches 3; Indels 0; Gaps 0;  
QY 1 MAVRALKLTLLAVVAASQAQSEAGAGWMTDPLLFAGTAAVARGDWGVLSMER 60  
Db 1 MAVRALKLTLLAVVAASQAQSEAGAGWMTDPLLFAGTAAVARGDWGVLSMER 60  
QY 61 ALRSRALRALRLRQTCQAADPWLDPSPAQASGAGALRDLSPFGGLRRAACL 120  
Db 61 ALRSRALRALRLRQTCQAADPWLDPSPAQASGAGALRDLSPFGGLRRAACL 120  
QY 121 RRLGLPPAAHSLSEMELEPRKSPYNYLQVAYFKINKLEKAVAAAHHTFVGNPEHMEMQ 180  
Db 121 RRLGLPPAAHSLSEMELEPRKSPYNYLQVAYFKINKLEKAVAAAHHTFVGNPEHMEMQ 180  
QY 181 QNLDYYQTMGKVEADFKDLETPHMOEFLGVRLYSESPQEAHPHLEAALQEVAYE 240  
Db 181 QNLDYYQTMGKVEADFKDLETPHMOEFLGVRLYSESPQEAHPHLEAALQEVAYE 240  
QY 241 ECRALCEGPDYDGYNYLYNADLFOAITDHYIQLVNLCKQNCVTELASHPSREKPPEDFL 300  
Db 241 ECRALCEGPDYDGYNYLYNADLFOAITDHYIQLVNLCKQNCVTELASHPSREKPPEDFL 300  
QY 301 PSHNYLQFAYNYGNQYTGACAKYLLFFPNDEVMQNLYAYAAMLGEHSTRSGPRE 360  
Db 301 PSHNYLQFAYNYGNQYTGACAKYLLFFPNDEVMQNLYAYAAMLGEHSTRSGPRE 360  
QY 361 SAKEXRQSLLEKELFFAYDVDFGIPDPSMTPEVTPKRLQEKOKSERETAVRISOE 420  
Db 361 SAKEXRQSLLEKELFFAYDVDFGIPDPSMTPEVTPKRLQEKOKSERETAVRISOE 420  
QY 421 IGLNLMKEIETLVEKTKESLDVSRKLTREGGPLYLLEGISLTMSKLANGYQVRVMDGVISD 480  
Db 421 IGLNLMKEIETLVEKTKESLDVSRKLTREGGPLYLLEGISLTMSKLANGYQVRVMDGVISD 480

QY 481 HECOELORLTNVAATSGDGYRGQTSPTPNKFKYGVTVFKALKLQSGKVPLOSAHLYYN 540  
Db 481 HECOELORLTNVAATSGDGYRGQTSPTPNKFKYGVTVFKALKLQSGKVPLOSAHLYYN 540  
QY 541 VTEKVRIMESYFRDLTPLYFSYSHLVCRTAIEVQAEKDDSHPHVDNCILNAETLVC 600  
Db 541 VTEKVRIMESYFRDLTPLYFSYSHLVCRTAIEVQAEKDDSHPHVDNCILNAETLVC 600  
QY 601 VKEPPAYTFRDYSAILYLNGDFDGNFYFTTDLAKTVTAEVQPCGCGRAVSGSGTENPHG 660  
Db 601 VKEPPAYTFRDYSAILYLNGDFDGNFYFTTDLAKTVTAEVQPCGCGRAVSGSGTENPHG 660  
QY 661 VKAVTRGRCALAWFTLDPHRSERDRVQADLVKMLFSPREMDLSOEQPLDAQGGPEP 720  
Db 661 VKAVTRGRCALAWFTLDPHRSERDRVQADLVKMLFSPREMDLSOEQPLDAQGGPEP 720  
QY 721 AQESLSGSESKPKDEL 736  
Db 721 AQESLSGSESKPKDEL 736  
RESULT 6  
AAB93142  
ID AAB93142 standard; protein; 736 AA.  
XX AAB93142;  
AC AAB93142;  
DT 26-JUN-2001 (first entry)  
XX Human protein sequence SEQ ID NO:12045.  
DE Human; primer; detection; diagnosis; antisense therapy; gene therapy.  
KW Homo sapiens.  
OS Homo sapiens.  
XX EP1074617-A2.  
PN 07-FEB-2001.  
PD 28-JUL-2000; 2000EP-00116126.  
PF 29-JUL-1999; 99JP-00248036.  
PR 27-AUG-1999; 99JP-00300253.  
PR 11-JAN-2000; 2000JP-00118776.  
PR 02-MAY-2000; 2000JP-00183767.  
PR 09-JUN-2000; 2000JP-00241899.  
XX (HELI-) HELIX RES INST.  
XX Ota T, Isogai T, Nishikawa T, Hayashi K, Saito K, Yamamoto J;  
PI Ishii S, Sugiyama T, Wakamatsu A, Nagai K, Otsuki T;  
XX WPI; 2001-318749/34.  
DR  
XX  
PT Primer sets for synthesizing polynucleotides, particularly the 5602 full-length cDNAs defined in the specification, and for the detection and/or diagnosis of the abnormality of the proteins encoded by the full-length cDNAs.  
PT  
XX  
PS Claim 8; SEQ ID NO 12045; 2537pp + Sequence Listing; English.  
XX  
CC The present invention describes primer sets for synthesising 5602 full-length cDNAs defined in the specification. Where a primer set comprises: (a) an oligo-dT primer and an oligonucleotide complementary to the complementary strand of a polynucleotide which comprises one of the 5602 nucleotide sequences defined in the specification, where the oligonucleotide comprises at least 15 nucleotides; or (b) a combination of an oligonucleotide comprising a sequence complementary to the complementary strand of a polynucleotide which comprises a 5'-end sequence and an oligonucleotide comprising a sequence complementary to a polynucleotide which comprises a 3'-end sequence, where the oligonucleotide comprises at least 15 nucleotides and the combination of

CC the 5'-end sequence/3'-end sequence is selected from those defined in the  
CC specification. The primer sets can be used in antisense therapy and in  
CC gene therapy. The primers are useful for synthesising polynucleotides,  
CC particularly full-length cDNAs. The primers are also useful for the  
CC detection and/or diagnosis of the abnormality of the proteins encoded by  
CC the full-length cDNAs. The primers allow obtaining of the full-length  
CC cDNAs easily without any specialised methods. AAH03166 to AAH13628 and  
CC AAH13633 to AAH18742 represent human cDNA sequences; AAB92446 to AAB95893  
CC represent human amino acid sequences; and AAH13629 to AAH13632 represent  
CC oligonucleotides, all of which are used in the exemplification of the  
CC present invention  
XX  
SQ Sequence 736 AA;

Query Match 99.3%; Score 3844; DB 4; Length 736;  
Best Local Similarity 99.5%; Pred. No. 0;  
Matches 732; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

Qy	1	MAVRALKLLTTLAVAAASQAQVESEAGMGWVTPDLLFAEGTAAYARGDWPQGVLSMER	60
Db	1	MAVRALKLLTTLAVAAASQAQVESEAGMGWVTPDLLFAEGTAAYARGDWPQGVLSMER	60
Qy	61	ALRSRAALRALRLCRTOCAADPWELDPWSPPAQASGAGALRDLSPFGGLRRAACL	120
Db	61	ALRSRAALRALRLCRTOCAADPWELDPWSPPAQASGAGALRDLSPFGGLRRAACL	120
Qy	121	RRCLGPPAAHSLSEMELEPRKSPNYLVQVFKINKLEKAVAAAHFTFVGNPEHMEQ	180
Db	121	RRCLGPPAAHSLSEMELEPRKSPNYLVQVFKINKLEKAVAAAHFTFVGNPEHMEQ	180
Qy	181	QNLDDYQTMGSKVKEADFKDLTPHMQEFLGRLVLYSEEQPOEAVPHLEALQEYVAYE	240
Db	181	QNLDDYQTMGSKVKEADFKDLTPHMQEFLGRLVLYSEEQPOEAVPHLEALQEYVAYE	240
Qy	241	ECRALCEGPDYDGYNLYEYNADLFOAITDHYIQVNLCKNCVTELASHPSREKPFEDFL	300
Db	241	ECRALCEGPDYDGYNLYEYNADLFOAITDHYIQVNLCKNCVTELASHPSREKPFEDFL	300
Qy	301	PSHNYLQFAYNIGNYQTOGCAKTYLLFPFNDVNMNQNLAYYAMLGEHTRSIGPRE	360
Db	301	PSHNYLQFAYNIGNYQTOGCAKTYLLFPFNDVNMNQNLAYYAMLGEHTRSIGPRE	360
Qy	361	SAKEYRORSLEKELLFPFVDVFGIPFVDPDSWTPPEVIPKRLQEKOKSERETAVRISOE	420
Db	361	SAKEYRORSLEKELLFPFVDVFGIPFVDPDSWTPPEVIPKRLQEKOKSERETAVRISOE	420
Qy	421	IGNLMKEIETLVBKTKESLDVSRLLTREGGPLLVEGISTLMSKLLNGYQVVMQGVISD	480
Db	421	IGNLMKEIETLVBKTKESLDVSRLLTREGGPLLVEGISTLMSKLLNGYQVVMQGVISD	480
Qy	481	HECOELQLTNVAATSGDVGRTSHTPNKEKFGVTVFKALKLGOEKGKVPLOSALYLYN	540
Db	481	HECOELQLTNVAATSGDVGRTSHTPNKEKFGVTVFKALKLGOEKGKVPLOSALYLYN	540
Qy	541	VTEKVRIMESYFRDLTPFLYFSYSHLVCRITAEVQAEKRDSDSHPVHVDNCILNAETLVC	600
Db	541	VTEKVRIMESYFRDLTPFLYFSYSHLVCRITAEVQAEKRDSDSHPVHVDNCILNAETLVC	600
Qy	601	VKEPPATVFRDYSNLLYNGDFGGNPFYFTELLDAKVTVAEVPQCGRVGFSSGTENPHG	660
Db	601	VKEPPATVFRDYSNLLYNGDFGGNPFYFTELLDAKVTVAEVPQCGRVGFSSGTENPHG	660
Qy	661	VKAVTRQRCALALFWFLDPRHSERDRVQADDLVKMLFSPSEMDLSQEQPLDAQQGFPEP	720
Db	661	VKAVTRQRCALALFWFLDPRHSERDRVQADDLVKMLFSPSEMDLSQEQPLDAQQGFPEP	720
Qy	721	AQESLSGSESKPKDEL 736	
Db	721	AQESLSGSESKPKDEL 736	

ID ABG66709 standard; protein; 806 AA.

XX ABG66709;

DT 30-AUG-2002 (first entry)

DE Human novel polypeptide #44.

KW Human; inflammatory condition; shock; sepsis; immune response; cancer;  
KW wound healing; central nervous system disease; haematopoiesis;  
KW peripheral nervous system disease; amyotrophic lateral sclerosis; tendon;  
KW myeloid cell disorder; lymphoid cell disorder; platelet disorder; bone;  
KW cartilage; ligament; nerve tissue; ulcer; osteoporosis; osteoarthritis;  
KW bone degenerative disorder; periodontal disease; reperfusion injury;  
KW lung fibrosis; liver fibrosis; autoimmune disorder; bacterial infection;  
KW allergic condition; thrombolysis; thrombosis; coagulation disorder;  
KW fungal infection.

XX Homo sapiens.

XX WO200244340-A2.

XX 06-JUN-2002.

XX 30-NOV-2001; 2001WO-US047004.

XX 30-NOV-2000; 2000US-00728952.

XX (HYSE-) HYSEQ INC.

XX Tang YT, Goodrich RW, Liu C, Zhou P, Asundi V, Wang J, Wang D;  
XX Yamazaki V, Ujwal ML, Drmanac RT;

XX WPI; 2002-508509/54.

XX N-PSDB; ABK94933.

XX Novel nucleic acids and polypeptides for diagnosis, treatment of  
XX inflammatory, autoimmune, nervous system, myeloid or lymphoid cell  
XX disorders, cancer and promoting wound healing.

XX Claim 10; Page 616-617; 672pp; English.

XX The invention relates to human novel polynucleotides and associated  
XX polypeptides. The polynucleotides and polypeptides are useful for  
XX treating inflammatory conditions such as arthritis, nephritis, Crohn's  
XX disease, ischaemia-reperfusion injury, shock, sepsis, immune responses  
XX and cancer and for promoting wound healing. The sequences are used to  
XX induce the proliferation of neural cells and regeneration of nerve and  
XX brain tissue, and are useful for the treatment of central and peripheral  
XX nervous system diseases and neuropathies, such as Alzheimer's disease,  
XX Parkinson's disease. Huntington's disease and amyotrophic lateral  
XX sclerosis. The sequences are involved in chemotactic or chemokinetic  
XX activity, regulation of haematopoiesis, treatment of myeloid or lymphoid  
XX cell disorders and platelet disorders such as thrombocytopenia,  
XX regeneration of bone, cartilage, tendon, ligament and/or nerve tissue  
XX growth, tissue repair, healing of burns, incisions, ulcers, treatment of  
XX osteoporosis, osteoarthritis, bone degenerative disorders and periodontal  
XX disease. The sequences of the invention are also useful for gut  
XX protection or regeneration and treatment of lung or liver fibrosis,  
XX reperfusion injury in various tissues, immune deficiencies and disorders  
XX including severe combined immunodeficiency (SCID), bacterial or fungal  
XX infections, autoimmune disorders e.g. multiple sclerosis and myasthenia  
XX gravis, allergic conditions such as asthma, thrombolysis or thrombosis  
XX and coagulation disorders. Sequences ABG66666-ABG66758 represent human  
XX novel polypeptides of the invention

XX Sequence 806 AA;

Query Match 98.2%; Score 3801; DB 5; Length 806;

Best Local Similarity 90.9%; Pred. No. 0;

Matches 733; Conservative 0; Mismatches 3; Indels 70; Gaps 1;

Qy 1 MAVRALKLLTTLAVAAASQAQVESEAGMGWVTPDLLFAEGTAAYARGDWPQGVLSMER 60

Db 1 MAVRALKLLTLLAVVAASQAQVESEAGWMTPTDLLFAEGTAAAYARGDWPVGLSMER 60  
Qy 61 ALRSRAALRALRLCRCTQCAADFPWELDPDWSPSQAQAGAGALRDLSPFGGLLRAACL 120  
Db 61 ALRSRAALRALRLCRCTQCAADFPWELDPDWSPSQAQAGAGALRDLSPFGGLLRAACL 120  
Qy 121 RRCGLPPAAHSLSEMELEFRKSPYNYLQVAYF-----KINKLEKAVAAAHFFF 154  
Db 121 RRCGLPPAAHSLSEMELEFRKSPYNYLQVAYF-----KINKLEKAVAAAHFFF 180  
Qy 155 -----KINKLEKAVAAAHFFF 170  
Db 181 GDRGVRREGKVASWLGSSPRSGELLPGRRPSPSSHGQMLTFKINKLEKAVAAAHFFF 240  
Qy 171 VGNPEHMEMQNLDYYQTMGSKVEADFKDLETPHMQEFLRGVRLYSEEQPQAVPHLEA 230  
Db 241 VGNPEHMEMQNLDYYQTMGSKVEADFKDLETPHMQEFLRGVRLYSEEQPQAVPHLEA 300  
Qy 231 ALQCYFYVAYBEACALCEGPDYDGYNYLEYNADLFOAITDHYIQVLNCKNCVCUTELASHP 290  
Db 301 ALQCYFYVAYBEACALCEGPDYDGYNYLEYNADLFOAITDHYIQVLNCKNCVCUTELASHP 360  
Qy 291 SREKPFEDFLPSHNYLQFYAYNIGNYTOAGECAKTYLLFFPNDEVMNQNLAYYAMLGE 350  
Db 361 SREKPFEDFLPSHNYLQFYAYNIGNYTOAGECAKTYLLFFPNDEVMNQNLAYYAMLGE 420  
Qy 351 EHTSICGPRSAKYRQSRSLLEKELLFPAYDVFQIPVDPSWTPPEVPIKRLQEKQKSE 410  
Db 421 EHTSICGPRSAKYRQSRSLLEKELLFPAYDVFQIPVDPSWTPPEVPIKRLQEKQKSE 480  
Qy 411 RETAVRISQETGNLMEKIEITLVEKTKESLDVSLRTREGGPLYEGISLTMNSKLNGYQ 470  
Db 481 RETAVRISQETGNLMEKIEITLVEKTKESLDVSLRTREGGPLYEGISLTMNSKLNGSQ 540  
Qy 471 RVVMDGVISDHECQELQRLTNVAATSGDGYRGQTSPTPNEKFGYVTVFKALXLBQEGKV 530  
Db 541 RVVMDGVISDHECQELQRLTNVAATSGDGYRGQTSPTPNEKFGYVTVFKALXLBQEGKV 600  
Qy 531 PLQSAHLYNVTXKVRIMESYFRDLTPLYFSYSHLVCRTAIEVQAEKRDSDHPVHVDN 590  
Db 601 PLQSAHLYNVTXKVRIMESYFRDLTPLYFSYSHLVCRTAIEVQAEKRDSDHPVHVDN 660  
Qy 591 CILNAETLVCVKEPPATFDYSAIYLNGDFDGNFYFTLDAKTVAEVQPCQGRAVG 650  
Db 661 CILNAETLVCVKEPPATFDYSAIYLNGDFDGNFYFTLDAKTVAEVQPCQGRAVG 720  
Qy 651 PSSGTENPHGVKAVTRGQRCALALWFTLDPHRSERDRVQADDLVKMLFSPSEMDLSQEQP 710  
Db 721 PSSGTENPHGVKAVTRGQRCALALWFTLDPHRSERDRVQADDLVKMLFSPSEMDLSQEQP 780  
Qy 711 LDAQGGPPEPAQESLSGSESKPKDEL 736  
Db 781 LDAQGGPPEPAQESLSGSESKPKDEL 806  
RESULT 8  
ID ABG66686 standard; protein; 806 AA.  
XX ABG66686;  
AC ABG66686;  
XX 30-AUG-2002 (first entry)  
XX Human novel polypeptide #21.  
XX Human; inflammatory condition; shock; sepsis; immune response; cancer;  
KW wound healing; central nervous system disease; hematopoiesis;  
KW peripheral nervous system disease; amyotrophic lateral sclerosis; tendon;  
KW myeloid cell disorder; lymphoid cell disorder; platelet disorder; bone;  
KW cartilage; ligament; nerve tissue; ulcer; osteoporosis; osteoarthritis;  
KW bone degenerative disorder; periodontal disease; reperfusion injury;  
KW lung fibrosis; liver fibrosis; autoimmune disorder; bacterial infection;

KW allergic condition; thrombolysis; thrombosis; coagulation disorder;  
KW fungal infection.  
XX Homo sapiens.  
PN WO200244340-A2.  
XX 06-JUN-2002.  
XX 30-NOV-2001; 2001WO-US047004.  
XX 30-NOV-2000; 2000US-00728952.  
XX (HYSE-) HYSEQ INC.  
XX Tang YT, Goodrich RW, Liu C, Zhou P, Asundi V, Wang J, Wang D;  
PI Yamazaki V, Ujwal ML, Drmanac RT;  
XX WPI: 2002-508509/54.  
DR N-PSDB; ABR94910.  
XX Novel nucleic acids and polypeptides for diagnosis, treatment of  
PT inflammatory, autoimmune, nervous system, myeloid or lymphoid cell  
PT disorders, cancer and promoting wound healing.  
XX Claim 10; Page 584-585; 672pp; English.  
XX The invention relates to human novel polynucleotides and associated  
CC polypeptides. The polynucleotides and polypeptides are useful for  
CC treating inflammatory conditions such as arthritis, nephritis, Crohn's  
CC disease, ischaemia-reperfusion injury, shock, sepsis, immune responses  
CC and cancer and for promoting wound healing. The sequences are used to  
CC induce the proliferation of neural cells and regeneration of nerve and  
CC brain tissue, and are useful for the treatment of central and peripheral  
CC nervous system diseases and neuropathies, such as Alzheimer's disease,  
CC Parkinson's disease, Huntington's disease and amyotrophic lateral  
CC sclerosis. The sequences are involved in chemotactic or chemokinetic  
CC activity, regulation of haematopoiesis, treatment of myeloid or lymphoid  
CC cell disorders and platelet disorders such as thrombocytopenia,  
CC regeneration of bone, cartilage, tendon, ligament and/or nerve tissue  
CC growth, tissue repair, healing of burns, incisions, ulcers, treatment of  
CC osteoporosis, osteoarthritis, bone degenerative disorders and periodontal  
CC disease. The sequences of the invention are also useful for gut  
CC protection or regeneration and treatment of lung or liver fibrosis,  
CC reperfusion injury in various tissues, immune deficiencies and disorders  
CC including severe combined immunodeficiency (SCID), bacterial or fungal  
CC infections, autoimmune disorders e.g. multiple sclerosis and myasthenia  
CC gravis, allergic conditions such as asthma, thrombolysis or thrombosis  
CC and coagulation disorders. Sequences ABG66666-ABG66758 represent human  
CC novel polypeptides of the invention  
XX  
SQ Sequence 806 AA;  
Query Match 98.2%; Score 3801; DB 5; Length 806;  
Best Local Similarity 90.9%; Pred. No. 0;  
Matches 733; Conservative 0; Mismatches 3; Indels 70; Gaps 1;

Qy 1 MAVRALKLLTLLAVVAASQAQVESEAGWMTPTDLLFAEGTAAAYARGDWPVGLSMER 60  
Db 1 MAVRALKLLTLLAVVAASQAQVESEAGWMTPTDLLFAEGTAAAYARGDWPVGLSMER 60  
Qy 61 ALRSRAALRALRLCRCTQCAADFPWELDPDWSPSQAQAGAGALRDLSPFGGLLRAACL 120  
Db 61 ALRSRAALRALRLCRCTQCAADFPWELDPDWSPSQAQAGAGALRDLSPFGGLLRAACL 120  
Qy 121 RRCGLPPAAHSLSEMELEFRKSPYNYLQVAYF-----KINKLEKAVAAAHFFF 154  
Db 121 RRCGLPPAAHSLSEMELEFRKSPYNYLQVAYF-----KINKLEKAVAAAHFFF 180  
Qy 155 -----KINKLEKAVAAAHFFF 170  
Db 181 GDRGVRREGKVASWLGSSPRSGELLPGRRPSPSSHGQMLTFKINKLEKAVAAAHFFF 240



QY 171 VGNPEHMEQNLDDYYQTMGSKVEADFKDLTQPHMQEFLRGVLVSEBQPEAVPHLEA 230  
 DB 241 VGNPEHMEQNLDDYYQTMGSKVEADFKDLTQPHMQEFLRGVLVSEBQPEAVPHLEA 300  
 QY 231 ALQSEYVAYEBCALCEGPDYDGYNYLYEYNADLFOAITDHYIOVLNCKQNCVTELASHP 290  
 DB 301 ALQSEYVAYEBCALCEGPDYDGYNYLYEYNADLFOAITDHYIOVLNCKQNCVTELASHP 360  
 QY 291 SREKPFDFLPSHNYLQFAYYINIGNYTOAGECAKTYLLFFPNDEVNQNLAAYAAIMGE 350  
 DB 361 SREKPFDFLPSHNYLQFAYYINIGNYTOAGECAKTYLLFFPNDEVNQNLAAYAAIMGE 420  
 QY 351 EHTRSIGPRESAKYRORSLEKELLFFAYDVFQIPVDPDSWTPPEVPIKRLQEKQKSE 410  
 DB 421 EHTRSIGPRESAKYRORSLEKELLFFAYDVFQIPVDPDSWTPPEVPIKRLQEKQKSE 480  
 QY 411 RETAVRISQELGNLKMIEITLVEKTKESLDVSRLTREGGPLLVEGSLTWNKLLNGYQ 470  
 DB 481 RETAVRISQELGNLKMIEITLVEKTKESLDVSRLTREGGPLLVEGSLTWNKLLNGSQ 540  
 QY 471 RVMDGVISDHECQELQRLTNVAATSGDGYRGQTSPTPNKFKYGVTVFVKALKLGQEGKV 530  
 DB 541 RVMDGVISDHECQELQRLTNVAATSGDGYRGQTSPTPNKFKYGVTVFVKALKLGQEGKV 600  
 QY 531 PLQSAHLYYNTVETKVRIMESYFRDLTPLYFSYSHLVCRTAIEVQAEKDDSHPVHVDN 590  
 DB 601 PLQSAHLYYNTVETKVRIMESYFRDLTPLYFSYSHLVCRTAIEVQAEKDDSHPVHVDN 660  
 QY 591 CILNAETLVCKEPPAYTFRDYSAILYNGDFDGNFYFTLDAKTVAEVPQCGRAVG 650  
 DB 661 CILNAETLVCKEPPAYTFRDYSAILYNGDFDGNFYFTLDAKTVAEVPQCGRAVG 720  
 QY 651 PSSGTENPHGVKATRGORCAIALWFTLDRHSERDRVQADDLVKMLFSPSEMDLSQEQP 710  
 DB 721 PSSGTENPHGVKATRGORCAIALWFTLDRHSERDRVQADDLVKMLFSPSEMDLSQEQP 780  
 QY 711 LDAQGPPEPAQESLSGSESKPKDEL 736  
 DB 781 LDAQGPPEPAQESLSGSESKPKDEL 806  
 RESULT 9  
 ID ABP69128  
 XX AC ABP69128;  
 XX DT 20-JAN-2003 (first entry)  
 XX DE Human polypeptide SEQ ID NO 1175.  
 KW Human: genome mapping; gene therapy; food supplement; virus; fungus;  
 KW cell-proliferative disorder; neurodegenerative disease; bacterial;  
 KW Parkinson's disease; Alzheimer's disease; autoimmune disease;  
 KW multiple sclerosis; diabetes; genetic disorder; wound; burn; infection;  
 KW arthritis; cytostatic; immunomodulator; neurotropic; dermatological;  
 KW antiparkinsonian; antidiabetic; immunosuppressive; hematological;  
 KW haemostatic; vulnerary; fungicide; antibacterial; virucide; protozoacide;  
 KW antiarthritic.  
 XX OS Homo sapiens.  
 XX FN WO200270539-A2.  
 XX PD 12-SEP-2002.  
 XX PF 05-MAR-2002; 2002WO-US005095.  
 XX PR 05-MAR-2001; 2001US-00799451.  
 XX PA (HYSE-) HYSEQ INC.  
 XX PI Tang YT, Zhou P, Goodrich RW, Asundi V, Zhang J, Zhao QA, Ren P;

PI Xue AJ, Yang Y, Ma Y, Yamazaki V, Chen R, Wang Z, Ghosh M;  
 PI Wehrman T, Wang J, Wang D, Drmanac RT;  
 XX WPI: 2002-759812/82.  
 DR N-PSDB; ABZ11345.  
 XX New polynucleotides comprising sequences assembled from expressed  
 PT sequence tags (ESTs), useful for treating cell-proliferative,  
 PT neurodegenerative, autoimmune, genetic, myeloid or lymphoid, or platelet  
 PT or coagulation disorders.  
 XX Claim 9; SEQ ID NO 1175; 1012pp + Sequence Listing; English.  
 PS The invention relates to an isolated polynucleotide (1) comprising a  
 XX nucleotide sequence selected from any of 948 sequences (ABZ11119-  
 CC ABZ12066) or their mature protein coding portion, active domain coding  
 CC protein or complementary sequences. The polynucleotides are useful for  
 CC identifying expressed genes or for physical mapping of human genome. The  
 CC encoded polypeptides (ABP68902-ABP69849) are useful as molecular weight  
 CC markers, as a food supplement, for generating antibodies, in medical  
 CC imaging, screening and diagnostic assays and for treating cell-  
 CC proliferative disorders (cancer), neurodegenerative diseases (Parkinson's  
 CC or Alzheimer's disease), autoimmune diseases (multiple sclerosis,  
 CC diabetes, lupus) genetic disorders, myeloid or lymphoid disorders,  
 CC platelet or coagulation disorders, wound, burns, incision, ulcers, liver  
 CC or lung fibrosis, infections (bacterial, viral, fungal, parasitic),  
 CC arthritis, etc. Note: the sequence data for this patent did not form part  
 CC of the printed specification, but was obtained in electronic format  
 CC directly from WIPO at ftp.wipo.int/pub/published\_pct\_sequences  
 XX Sequence 706 AA;  
 SQ  
 Query Match 94.2%; Score 3645; DB 5; Length 706;  
 Best Local Similarity 95.0%; Pred. No. 1.1e-312;  
 Matches 699; Conservative 2; Mismatches 5; Indels 30; Gaps 1;  
 QY 1 MAVRALKLTITLAVVAAASQAEVSEAGWGWITPDLFAEGTAAAYARGDWGVLSNER 60  
 DB 1 MAVRALKLTITLAVVAAASQAEVSEAGWGWITPDLFAEGTAAAYARGDWGVLSNER 60  
 QY 61 ALRSRAALRALRLRTOCAADFPWELDPDWSPPAQASGAGALRDLISFFGLLRAACL 120  
 DB 61 ALRSRAALRALRLRTOCAADFPWELDPDWSPPAQASGAGALRDLISFFGLLRAACL 120  
 QY 121 RRLCLGPPAAHSLSEEMELEFRKSPYNYLVQAYFKINKLEKAVAAAHFTFFVGNPEHMEQ 180  
 DB 121 RRLCLGPPAAHSLSEEMELEFRKSPYNYLVQAYFKINKLEKAVAAAHFTFFVGNPEHMEQ 180  
 QY 181 QNLDDYYQTMGSKVEADFKDLTQPHMQEFLRGVLVSEBQPEAVPHLEAALQYFVAYE 240  
 DB 181 QNLDDYYQTMGSKVEADFKDLTQPHMQEFLRGVLVSEBQPEAVPHLEAALQYFVAYE 240  
 QY 241 ECRALCEGPDYDGYNYLYEYNADLFOAITDHYIOVLNCKQNCVTELASHPSREKPFDFDL 300  
 DB 241 ECRALCEGPDYDGYNYLYEYNADLFOAITDHYIOVLNCKQNCVTELASHPSREKPFDFDL 300  
 QY 301 PSHYNYLQFAYYINIGNYTOAGECAKTYLLFFPNDEVNQNLAAYAAIMGEHTSISIGRE 360  
 DB 301 PSHYNYLQFAYYINIGNYTOAGECAKTYLLFFPNDEVNQNLAAYAAIMGEHTSISIGRE 360  
 QY 361 SAKYRORSLEKELLFFAYDVFQIPVDPDSWTPPEVPIKRLQEKQKSERETAVRISQ 420  
 DB 361 SAKYRORSLEKELLFFAYDVFQIPVDPDSWTPPEVPIKRLQEKQKSERETAVRISQ 420  
 QY 421 IGLNLMKEIETLVEKTKESLDVSRLTREGGPLLVEGSLTWNKLLNGYQRVVMDGVISD 480  
 DB 391 IGLNLMKEIETLVEKTKESLDVSRLTREGGPLLVEGSLTWNKLLNGYQRVVMDGVISD 450  
 QY 481 HECQELQRLTNVAATSGDGYRGQTSPTPNKFKYGVTVFVKALKLGQEGKVPLQSAHLYN 540  
 DB 451 HOCQELQRLTNVAATSGDGYRGQTSPTPNKFKYGVTVFVKALKLGQEGKVPLQSAHLYN 510  
 QY 541 VTEKVRIMESYFRDLTPLYFSYSHLVCRTAIEVQAEKDDSHPVHVDNLCILNAETLV 600



511 VTEKRRIMESYFRLDTPLYFSYSHLVCRITAEVQAERKDDSHPVHVDNCILNAETLVC 570  
601 VKPPPAYTFRDYSAIILYLNGDFGCGHFFFTLDAKTVTAHVQPCGRVGFSSGTENPHG 660  
571 VKEPPPAYTFRDYSAIILYLNGDFGCGHFFFTLDAKTVTAHVQPCGRVGFSSGTENPHG 630  
661 VKAVTRGORCAIALWFTLDRHSERDVRQADDLVXKMLFSPPEMDLSQEQPLDQAQQPPPP 720  
631 VKAVTRGORCAIALWFTLDRHSERDVRQADDLVXKMLFSPPEMDLSQEQPLDQAQQPPPP 690  
721 AQESLSGSSEKPKDEL 736  
691 AQESLSGSSEKPKDEL 706

RESULT 10  
AAB94482  
ID AAB94482 standard; protein; 804 AA.  
XX  
AC AAB94482;  
XX  
DT 26-JUN-2001 (first entry)  
XX  
DE Human protein sequence SEQ ID NO:15160.  
XX  
KW Human; primer; detection; diagnosis; antisense therapy; gene therapy.  
XX  
OS Homo sapiens.  
XX  
PN EF1074617-A2.  
XX  
PD 07-FEB-2001.  
XX  
PF 28-JUL-2000; 2000EP-00116126.  
XX  
PR 29-JUL-1999; 99JP-00248036.  
PR 27-AUG-1999; 99JP-00300253.  
PR 11-JAN-2000; 2000JP-00118776.  
PR 02-MAY-2000; 2000JP-00183767.  
PR 09-JUN-2000; 2000JP-00241899.  
XX  
PA (HELI-) HELIX RES INST.  
XX  
PI Ota T, Isogai T, Nishikawa T, Hayashi K, Saito K, Yamamoto J;  
PI Ishii S, Sugiyama T, Wakamatsu A, Nagai K, Otsuki T;  
XX  
DR WPI; 2001-318749/34.  
XX  
PT Primer sets for synthesizing polynucleotides, particularly the 5602 full-length cDNAs defined in the specification, and for the detection and/or diagnosis of the abnormality of the proteins encoded by the full-length cDNAs.  
XX  
PS Claim 8; SEQ ID NO 15160; 2537pp + Sequence Listing; English.  
XX  
CC The present invention describes primer sets for synthesizing 5602 full-length cDNAs defined in the specification. Where a primer set comprises: (a) an oligo-dr primer and an oligonucleotide complementary to the complementary strand of a polynucleotide which comprises one of the 5602 nucleotide sequences defined in the specification, where the oligonucleotide comprises at least 15 nucleotides; or (b) a combination of an oligonucleotide comprising a sequence complementary to the complementary strand of a polynucleotide which comprises a 5'-end sequence and an oligonucleotide comprising a sequence complementary to a polynucleotide which comprises a 3'-end sequence, where the oligonucleotide comprises at least 15 nucleotides and the combination of the 5'-end sequence/3'-end sequence is selected from those defined in the specification. The primer sets can be used in antisense therapy and in gene therapy. The primers are useful for synthesizing polynucleotides, particularly full-length cDNAs. The primers are also useful for the detection and/or diagnosis of the abnormality of the proteins encoded by the full-length cDNAs. The primers allow obtaining of the full-length

CC cDNAs easily without any specialised methods. AAH03166 to AAH13628 and AAH13633 to AAH18742 represent human cDNA sequences; AAB92446 to AAB95893 represent human amino acid sequences; and AAH13629 to AAH13632 represent oligonucleotides, all of which are used in the exemplification of the present invention  
XX  
SQ Sequence 804 AA;  
Query Match 92.6%; Score 3582; DB 4; Length 804;  
Best Local Similarity 99.3%; Pred. No. 5.2e-307;  
Matches 682; Conservative 0; Mismatches 5; Indels 0; Gaps 0;  
QY 1 MAVRALKLLTLLAVVAAASQAEESEAGWGVTPTDLLFABGTAYAYAGDWPVGVLSMER 60  
DB 1 MAVRALKLLTLLAVVAAASQAEESEAGWGVTPTDLLFABGTAYAYAGDWPVGVLSMER 60  
QY 61 ALRSRAALRALRLRCRTQCAADFPEWLPDPSPPAQASGAGALRDLISFFGGLLRAACL 120  
DB 61 ALRSRAALRALRLRCRTQCAADFPEWLPDPSPPAQASGAGALRDLISFFGGLLRAACL 120  
QY 121 RRLGPPAAHLSLEMELEFRKRSYNYLQVAYFKINKLEKAVAAAHFTFFVGNPEHMEMQ 180  
DB 121 RRLGPPAAHLSLEMELEFRKRSYNYLQVAYFKINKLEKAVAAAHFTFFVGNPEHMEMQ 180  
QY 181 QNLDYYQTMGVKEADFKDLETPHMQBFRGLGVRLYSEEQPOEAVPHLEAALQYFYAYE 240  
DB 181 QNLDYYQTMGVKEADFKDLETPHMQBFRGLGVRLYSEEQPOEAVPHLEAALQYFYAYE 240  
QY 241 ECRALCEGPDYDGYNYLYNADLFOAITDHYIOVLNCKQNCVTELASHPSREKPFEDFL 300  
DB 241 ECRALCEGPDYDGYNYLYNADLFOAITDHYIOVLNCKQNCVTELASHPSREKPFEDFL 300  
QY 301 PSHYNYLOFAYYNYGNYTQAGECAKTYLLFFPNDEVNQNLAYYAAMLGEEHTRSIGPRE 360  
DB 301 PSHYNYLOFAYYNYGNYTQAGECAKTYLLFFPNDEVNQNLAYYAAMLGEEHTRSIGPRE 360  
QY 361 SAKYRQORSLLEKELLFPAYDVFGIPFVDPDSWTPPEVPIKRLQEKQKSERETAVRISQE 420  
DB 361 SAKYRQORSLLEKELLFPAYDVFGIPFVDPDSWTPPEVPIKRLQEKQKSERETAVRISQE 420  
QY 421 IGLNLMKEIETLVEEKTKESSLVSRITREGGPLLVEGILSTWNSKLINGYQVWMDGVISD 480  
DB 421 IGLNLMKEIETLVEEKTKESSLVSRITREGGPLLVEGILSTWNSKLINGYQVWMDGVISD 480  
QY 481 HECQELQRLTNVAATSGDGYRGQTSPTPNKPKFYGVTVFKALKLQEGKVPLOSAHLYYN 540  
DB 481 HECQELQRLTNVAATSGDGYRGQTSPTPNKPKFYGVTVFKALKLQEGKVPLOSAHLYYN 540  
QY 541 VTEKVRRIWESYFRDLTPLYFSYSHLVCRITAEVQAERKDDSHPVHVDNCILNAETLVC 600  
DB 541 VTEKVRRIWESYFRDLTPLYFSYSHLVCRITAEVQAERKDDSHPVHVDNCILNAETLVC 600  
QY 601 VKEPPPAYTFRDYSAIILYLNGDFGCGHFFFTLDAKTVTAHVQPCGRVGFSSGTENPHG 660  
DB 601 VKEPPPAYTFRDYSAIILYLNGDFGCGHFFFTLDAKTVTAHVQPCGRVGFSSGTENPHG 660  
QY 661 VKAVTRGORCAIALWFTLDRHSERDR 687  
DB 661 VKAVTRGORCAIALWFTLDRHSERDR 687

RESULT 11  
AAU09860  
ID AAU09860 standard; protein; 708 AA.  
XX  
AC AAU09860;  
XX  
DT 26-FEB-2002 (first entry)  
XX  
DE Novel human secreted protein #1.  
XX  
KW Secreted protein; cytostatic; immunosuppressive; vulnerary; vaccine; antiinflammatory; neuroprotective; nephrotropic; cardiovascular; human;  
KW

KW cancer; autoimmune disease; wound healing disorder; infection;  
 KW haematopoietic disorder; inflammatory disorder; infertility;  
 KW neurological disease; psychiatric disease; cardiovascular disease;  
 KW respiratory disease; renal; gastrointestinal.

XX Homo sapiens.

PN WO200179454-A1.

XX 25-OCT-2001.

XX 11-APR-2001; 2001WO-US011797.

XX 13-APR-2000; 2000US-0196603P.

PR 24-APR-2000; 2000US-0199417P.

XX (SMIK ) SMITHKLINE BEECHAM CORP.

PA (SMIK ) SMITHKLINE BEECHAM PLC.

DI Agarwal P, Murdoch PR, Rizvi SK, Smith RF, Xiang Z;

XX WPI; 2002-061975/08.

DR N-PSDB; AAS17572.

XX New secreted proteins or polypeptides, useful for treating e.g. cancer,  
 PT autoimmune diseases, wound healing disorder, infections, haematopoietic  
 PT disorders, inflammatory disorders, infertility, cancer.

XX Claim 1; Page 52-54; 92pp; English.

XX The invention relates to an isolated novel secreted polypeptide (I) and  
 CC polynucleotide (II). (I) and (II) are useful for treating cancer,  
 CC autoimmune diseases, wound healing disorder, infections, haematopoietic  
 CC disorders, inflammatory disorders, infertility, neurological and  
 CC psychiatric diseases, cardiovascular diseases, respiratory diseases,  
 CC renal diseases, or gastrointestinal diseases. These may also be used to  
 CC treat diseases, abnormalities and disorders caused by abnormal  
 CC expression, production, function and/or metabolism of the genes, as  
 CC vaccines for inducing immunological response in a mammal, and in  
 CC screening methods for detecting the effect of added compounds on the  
 CC production of mRNA and polypeptide in cells. The polypeptides can be used  
 CC as immunogens to produce antibodies immunospecific for the polypeptides,  
 CC and to identify membrane-bound or soluble receptors. The polynucleotides  
 CC may be used as diagnostic reagents, in chromosome localisation studies,  
 CC and in tissue expression studies. The present sequence represents the  
 CC amino acid sequence of novel human secreted protein #1

XX Sequence 708 AA;

Query Match 92.3%; Score 3572; DB 5; Length 708;  
 Best Local Similarity 93.9%; Pred. No. 3.2e-306;  
 Matches 691; Conservative 5; Mismatches 12; Indels 28; Gaps 3;

QY 1 MAVRALKLTLLAVAAASQAQVESBAGWGVTPTDLLFABGTAAYARGDWPVGWLSMER 60  
 DB 1 MAVRALKLTLLAVAAASQAQVESBAGWGVTPTDLLFABGTAAYARGDWPVGWLSMER 60  
 QY 61 ALRSGRAALRALRLCRTOCAADFFWELDPDWSQAQAGALRDLISFFGGLLRAACL 120  
 DB 61 ALRSGRAALRALRLCRTOCAADFFWELDPDWSQAQAGALRDLISFFGGLLRAACL 120  
 QY 121 RRCIGPPAAHSLSEMELEFRKSPYNYLOVAFKINKLEKAVAAAHFTFVGVPENMEMQ 180  
 DB 121 RRCIGPPAAHSLSEMELEFRKSPYNYLOVAFKINKLEKAVAAAHFTFVGVPENMEMQ 180  
 QY 181 QNLDDYQTMGSKVEADFKDLTQPHMQEFLRGVRLYSEEQPQAVPHLEALQEFVAYE 240  
 DB 181 QNLDDYQTMGSKVEADFKDLTQPHMQEFLRGVRLYSEEQPQAVPHLEALQEFVAYE 240  
 QY 241 ECRALCGPYDYGNYLYENADLFQAITHYIQVLNCKNCVTELASHPSREKPEDFL 300  
 DB 241 ECRALCGPYDYGNYLYENADLFQAITHYIQVLNCKNCVTELASHPSREKPEDFL 300

QY 301 PSHYNYLOFAYYNYGNVYQAGECAKTYLLFPFNDEVMNQNLAYYAAMLGBEHTRSIGPRE 360  
 DB 301 PSHYNYLOFAYYNYGNVYQAGECAKTYLLFPFNDEVMNQNLAYYAAMLGBEHTRSIGPRE 360  
 QY 361 SAKYRQSRSLLEKELLFFAYDVFGIPFVDPDSWTFPEEVIKRLQEKQKSERETAVRISQE 420  
 DB 333 SAKYRQSRSLLEKELLFFAYDVFGIPFVDPDSWTFPEEVIKRLQEKQKSERETAVRISQE 392  
 QY 421 IGNLMKEIETLVEEKTESLVSRLTREGGPLLLEGISLTWNSKLNGYQVWMDGVISD 480  
 DB 393 IGNLMKEIETLVEEKTESLVSRLTREGGPLLLEGISLTWNSKLNGYQVWMDGVISD 452  
 QY 481 HECQELQRLTNVAATSGDYGRTSPHTPNEKFGVTTFKALKLGQEGKVPLOSAHLYYN 540  
 DB 453 HECQELQRLTNVAATSGDYGRTSPHTPNEKFGVTTFKALKLGQEGKVPLOSAHLYYN 512  
 QY 541 VTEKVRIMESYFRDLTPLYFSYSHLVCRTAIEVQAEKDDSHPHVNDNCILNAETLVC 600  
 DB 513 VTEKVRIMESYFRDLTPLYFSYSHLVCRTAIEVQAEKDDSHPHVNDNCILNAETLVC 572  
 QY 601 VKEPPAYTFRDYSAILYLNGDFDGNFVFTELDAAVTAEVQPCGQRAVGFSSGTENPHG 660  
 DB 573 VKEPPAYTFRDYSAILYLNGDFDGNFVFTELDAAVTAEVQPCGQRAVGFSSGTENPHG 632  
 QY 661 VKAVTRGQRCALWFTLDPHRSERDRVQADDLVKMLFSPPEMDLSQBPDLDAQGPPEP 720  
 DB 633 VKAVTRGQRCALWFTLDPHRSERDRVQADDLVKMLFSPPEMDLSQBPDLDAQGPPEP 692  
 QY 721 AQESLSGSSEKPKDEL 736  
 DB 693 AQESLSGSSEKPKDEL 708

# RESULT 12

AAB36393

ID AAB36393 standard; protein; 747 AA.

AC AAB36393;

XX 27-FEB-2001 (first entry)

XX Mouse tumour suppressor Gros1-L protein SEQ ID NO:6.

XX Tumour suppressor; Gros1-L; Gros1-S; cell proliferation; regulation;  
 KW cancer; cytostatic; gene therapy.

XX Mus musculus.

XX WO200065047-A1.

XX 02-NOV-2000.

XX 26-APR-2000; 2000WO-JP002731.

XX 26-APR-1999; 99JP-00118806.

XX (CHUG-) CHUGAI RES INST MOLECULAR MEDICINE INC.

XX Wadhwa R, Sugihara T, Yoshida A;

XX WPI; 2000-687340/67.

XX N-PSDB; AAC64726.

XX Full-length tumor suppressor genes encoding Gros1-L, Gros1-S and mouse  
 PT homologs participating in regulation of cell proliferation, useful in  
 PT development of preventives and remedies of cancer.

XX Claim 1; Page 91-97; 114pp; Japanese.

XX The present sequence represents the mouse tumour suppressor designated  
 CC Gros1-L. Gros1-L and Gros1-S have cytostatic activity and can be used in  
 CC gene therapy. Gros1-L and Gros1-S genes are useful in the development of  
 CC drugs used to treat and prevent cancer

```
XX SQ Sequence 747 AA;
Query Match 85.7%; Score 3316; DB 3; Length 747;
Best Local Similarity 88.2%; Pred. No. 1.5e-283;
Matches 623; Conservative 33; Mismatches 48; Indels 2; Gaps 1;
QY 25 ESAGAGMVTPLDFAEGTAAYARGDWPVGLSVRAALRRLRCRTQCAADFP 84
Db 25 ESEPGDVAAPDLLYAEGTAAYSRRDWPVGLSVRAALRRLRCRTQCAADFP 84
QY 85 WELDPDWSPP--AQASGAGALRDLFFGGLLRAACLRCLGPPAAHSLSEMELEFRK 142
Db 85 WAPDLGLDPSLSQDPGAAALHDLFFGAVLRAACLRCLGPPAAHSLSEMELEFNK 144
QY 143 RSPYNYLOVAFKINKLEKAVAAAHFTFVGNPHEMEOQNLDYYQTMGKADFKDLET 202
Db 145 RSPYNYLOVAFKINKLEKAVAAAHFTFVGNPHEMEOQNLDYYQTMGKADFKDLEA 204
QY 203 QPHMOEFLGRLVLYSEEQPQEAAPHLEAALQEFVAYEBCRALCEGPDYDGYNYLEYNA 262
Db 205 KPHMHEFLGRLVLYSEBKQEAAPHLEAALQEFVAYEBCRALCEGPDYDGYNYLDYSA 264
QY 263 DLFOAITDHYIOVLNCKQNCVTELASHPSEKPFEDFLPSHYNLYQFAYYNGNTQAGE 322
Db 265 DLFOAITDHYIOVLNCKQNCVTELASHPSEKPFEDFLPSHYNLYQFAYYNGNTQAE 324
QY 323 CAKTYLLFFPNDEVMQNLAAYAAMLGHEHTRSIGPRESAKERYRQSLLEKELFFAYDV 382
Db 325 CAKTYLLFFPNDEVMQNLAAYAAMLGHEHTRSIGPRESAKERYRQSLLEKELFFAYDI 384
QY 383 FGIPFVDPDSWTPEEVIKRLQEKQKSERETAVRISQBIGNLKMEIETLVSEKTESLDV 442
Db 385 FGIPFVDPDSWTPEEVIKRLQEKQKSERETAVRISQBIGNLKMEIETLVSEKTESLDV 444
QY 443 SLRTREGGPLYEGISLTNWSKLVNGYQVMDGVISDHECOELQRLTNVAATSGDGYRG 502
Db 445 SLRTREGGPLYEGISLTNWSKLVNGYQVMDGVISDHECOELQRLTNVAATSGDGYRG 504
QY 503 QTSPTPNEKFGYVTVLKALQSGKVPLOSAHLYNNVTEKVRIMESYFRLDTPLYFS 562
Db 505 QTSPTPNEKFGYVTVLKALQSGKVPLOSAHLYNNVTEKVRIMESYFRLDTPLYFS 564
QY 563 YSHLVCEATAEVQAEKQDHPVVDNCTLNAETLVCKVEPPAYTFRDYSALLYNGDF 622
Db 565 YSHFVCKTAEESQAEKQDHPVVDNCTLNAETLVCKVEPPAYTFRDYSALLYNGDF 624
QY 623 DGGNFYFTELDAKTVAEVQPCQRAVGFSSGTENPHGVKAVTRGQRCALWFTLDPRH 682
Db 625 DGGNFYFTELDAKTVAEVQPCQRAVGFSSGTENPHGVKAVTRGQRCALWFTLDPRH 684
QY 683 SERDRVQADLVKMLFSPEWMDLSQEQPLDAQQGPPEPAQESLSGS 728
Db 685 SERDRVQADLVKMLFSPEWMDLPQEQPLPDQQGSPEPGEFLHGA 730
RESULT 13
ABR63160
ID ABR63160 standard; protein; 747 AA.
XX AC
XX AC ABR63160;
XX DT 18-DEC-2003 (first entry)
XX DE Murine polypeptide.
XX KW Mouse; prollyhydroxylase-like protein; enzyme; cyrostatic; cardiant;
XX KW antiaethmatic; neuroleptic; antiparkinsonian; neuroprotective; nootropic;
XX KW uropathic; gene therapy.
XX OS Mus musculus.
XX FN W0200306862-A1.
```

```
XX PD 14-AUG-2003.
XX PF 04-FEB-2003; 2003WO-EP001082.
XX PF 04-FEB-2002; 2002US-0353190P.
PR 09-MAY-2002; 2002US-0378674P.
XX PA (FARB ) BAYER AG.
XX PI Smith TJ;
XX XX WPI; 2003-663597/62.
DR WPI; 2003-663597/62.
XX XX New polynucleotide and its encoded prollyhydroxylase-like protein useful
PT for identifying modulators of protein activity and in gene therapy for
PT treating disorders, e.g. cancer, Alzheimer's disease and genitourinary
PT disorders.
XX PS Disclosure; Page 121-123; 135pp; English.
XX XX The present sequence is the protein sequence of a murine polypeptide. The
CC invention relates to human prollyhydroxylase-like protein (see ABR63159),
CC which is expressed in microvascular endothelial cells, prostate (normal
CC and cancerous) and colon epithelium. Human prollyhydroxylase-like protein
CC polynucleotides and polypeptides, agonists and antagonists can be used in
CC the prevention, amelioration or correction of dysfunctions or diseases
CC including cancer, cardiovascular disorders, chronic obstructive pulmonary
CC disease, asthma, genitourinary disorders, and central nervous system
CC disorders (claimed), e.g. schizophrenia, Alzheimer's disease and
CC Parkinson's disease
XX SQ Sequence 747 AA;
Query Match 85.7%; Score 3316; DB 7; Length 747;
Best Local Similarity 88.2%; Pred. No. 1.5e-283;
Matches 623; Conservative 33; Mismatches 48; Indels 2; Gaps 1;
```

QY 563 YSHLVCRFAIEVQAEKDDSHPHVDMNCILNAETLVCKVKEPPAYTRDYSAILYLNGDF 622  
 DB 565 YSHFVCRTAIESQAEKDDSHPHVDMNCILNAEAFMCIEKPEPPAYTRFYSAILYLNGDF 624  
 QY 623 DGGNFYFTELDAKTVAEVPQCGRAVGFSSGTENPHGVKAVTRGQRCALALWFTLDPRH 682  
 DB 625 DGGNFYFTELDAKTVAEVPQCGRAVGFSSGTENPHGVKAVTRGQRCALALWFTLDPRH 684  
 QY 683 SERDRVQADDLVXKMLFSPPEMDLQEQPLDAQQPPPPAQSLSGS 728  
 DB 685 SERDRVQADDLVXKMLFSPPEMDLQEQPLDAQQPPPPGEBFLHGA 730

# RESULT 14 AAE04244

ID AAE04244 standard; protein; 592 AA.

AC AAE04244;

DT 09-AUG-2001 (first entry)

DE Human gene 2 encoded secreted protein fragment, SEQ ID NO:100.

XX Human; secreted protein; proliferative disorder; cancer; tumour;  
 KW foetal abnormality; developmental abnormality; haematopoietic disorder;  
 KW immune system disorder; AIDS; autoimmune disease; rheumatoid arthritis;  
 KW inflammation; allergy; neurological disorder; Alzheimer's disease;  
 KW Parkinson's disease; cognitive disorder; schizophrenia; asthma;  
 KW skin disorder; psoriasis; sepsis; diabetes; atherosclerosis;  
 KW cardiovascular disorder; angiogenic disorder; kidney disorder;  
 KW gastrointestinal disorder; pregnancy-related disorder; gene therapy;  
 KW endocrine disorder; infection; wound healing; vulnetary; cell culture;  
 KW chemotaxis; food additive; binding partner identification.

XX Homo sapiens.

XX WO200136432-A2.

PD 25-MAY-2001.

PF 15-NOV-2000; 2000WO-US031162.

PR 19-NOV-1999; 99US-0166415P.

PR 30-JUN-2000; 2000US-0215136P.

XX (HUMA-) HUMAN GENOME SCI INC.

PI Ruben SM, Komatsoulis GA, Baker KP, Young PE;

DR WPI; 2001-343793/36.

PT Isolated nucleic acid molecule encoding a human secreted protein is used  
 in preventing, treating or ameliorating a medical condition.

PS Disclosure; Page 470-472; 509pp; English.

XX AAD08488-AAD08529 represent cDNAs corresponding to 18 human secreted  
 protein genes, and AAE04199-AAE04239 represent the proteins they encode.  
 CC AAE04240-AAE04297 represent human secreted protein fragments or variants.  
 CC The secreted proteins and their genes are useful for preventing, treating  
 CC or ameliorating medical conditions, e.g., by protein or gene therapy.  
 CC Pathological conditions can be diagnosed by determining the amount of the  
 CC new protein in a sample or by determining the presence of mutations in  
 CC the new genes. Specific uses are described for each of the 18 genes,  
 CC based on the tissues in which they are most highly expressed, and include  
 CC developing products for the diagnosis or treatment of proliferative  
 CC disorders, cancer, tumours, foetal and developmental abnormalities,  
 CC haematopoietic disorders, diseases of the immune system, AIDS, autoimmune  
 CC diseases (e.g., rheumatoid arthritis), inflammation, allergies, disease),  
 CC neurological disorders (e.g., Alzheimer's disease, Parkinson's disease),  
 CC cognitive disorders, schizophrenia, asthma, skin disorders (e.g.,  
 CC psoriasis), sepsis, diabetes, atherosclerosis, cardiovascular disorders,

CC angiogenic disorders, kidney disorders, gastrointestinal disorders,  
 CC pregnancy-related disorders, endocrine disorders, and infections. The  
 CC proteins can also be used to aid wound healing and epithelial cell  
 CC proliferation, to prevent skin ageing due to sunburn, to maintain organs  
 CC before transplantation, for supporting cell culture of primary tissues,  
 CC to regenerate tissues, to identify their cognate ligands or binding  
 CC partners, and in chemotaxis, and can be used as a food additive or  
 CC preservative to modify storage properties. Antibodies specific for a  
 CC protein of the invention can be used in alleviating symptoms associated  
 CC with the disorders mentioned above, and in diagnostic immunoassays e.g.,  
 CC radioimmunoassay or enzyme linked immunosorbent assay (ELISA). The  
 CC present sequence represents a human secreted protein fragment referred to  
 CC in the disclosure of the invention

XX Sequence 592 AA;

Query Match 80.5%; Score 3115; DB 4; Length 592;

Best Local Similarity 99.7%; Pred. No. 6.1e-266;

Matches 590; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 145 PYNLYQVAFKINKLEKAVAAHTFFVGNPEHMEMQQLNDYYQTMGSKYKADFKDLETQP 204

DB 1 PYNLYQVAFKINKLEKAVAAHTFFVGNPEHMEMQQLNDYYQTMGSKYKADFKDLETQP 60

QY 205 HMQEFLGLVRLYSEBPQOEAVPHLEAALQEVFVAYEECRALCEGPDYDGYNLYFNADL 264

DB 61 HMQEFLGLVRLYSEBPQOEAVPHLEAALQEVFVAYEECRALCEGPDYDGYNLYFNADL 120

QY 265 FOAITDHYIQLNQCNCVTELASHPSREKPEPFLPSHYNYLOPAYNIGNYQAGECA 324

DB 121 FOAITDHYIQLNQCNCVTELASHPSREKPEPFLPSHYNYLOPAYNIGNYQAGECA 180

QY 325 KYLLFFPNDEVMNQNLAYYAAMLGEETHRSIGPRESAKYRQSLLEKELLFFAYDVFG 384

DB 181 KYLLFFPNDEVMNQNLAYYAAMLGEETHRSIGPRESAKYRQSLLEKELLFFAYDVFG 240

QY 385 IPFVDPDSWTPEEVIKSLQEKQKSERETAVRISOEIGNLMKKEITLVEEKTESLDVSR 444

DB 241 IPFVDPDSWTPEEVIKSLQEKQKSERETAVRISOEIGNLMKKEITLVEEKTESLDVSR 300

QY 445 LTREGGPLLYEGISLTMNSKLINGQYVMDGVISDHCEQELQRLTNVAATSGGGRGQT 504

DB 301 LTREGGPLLYEGISLTMNSKLINGQYVMDGVISDHCEQELQRLTNVAATSGGGRGQT 360

QY 505 SPHTPNEKFGYVTVPFKALKQEGKVPLOQSAHLYNVTKEVRRIMESYFRDLTPLYFSYS 564

DB 361 SPHTPNEKFGYVTVPFKALKQEGKVPLOQSAHLYNVTKEVRRIMESYFRDLTPLYFSYS 420

QY 565 HLVCRTAIEEVOAEKDDSHPHVDMNCILNAETLVCKVKEPPAYTRDYSAILYLNGDPDG 624

DB 421 HLVCRTAIEEVOAEKDDSHPHVDMNCILNAETLVCKVKEPPAYTRDYSAILYLNGDPDG 480

QY 625 GNIFYTELDKATVTAEVQPCGQRAVGFSSGTENPHGVKAVTRGQRCALALWFTLDPHSE 684

DB 481 GNIFYTELDKATVTAEVQPCGQRAVGFSSGTENPHGVKAVTRGQRCALALWFTLDPHSE 540

QY 685 RDRVQADDLVXKMLFSPPEMDLSQEQPLDAQQPPPPAQSLSGSSEKPKDEL 736

DB 541 RDRVQADDLVXKMLFSPPEMDLSQEQPLDAQQPPPPAQSLSGSSEKPKDEL 592

## RESULT 15

XX AAB36394

ID AAB36394 standard; protein; 542 AA.

XX AAB36394;

DT 27-FEB-2001 (first entry)

DE Mouse tumour suppressor Gros1-S protein SEQ ID NO:8.

XX Tumour suppressor; Gros1-L; Gros1-S; cell proliferation; regulation;  
 KW cancer; cytostatic; gene therapy.

Job time : 110 secs

```
XX Mus musculus.
OS
XX
XX PN WO200065047-A1.
XX
XX PD
XX PF 02-NOV-2000.
XX
XX PR 26-APR-2000; 2000MO-JP002731.
XX
XX PR 26-APR-1999; 99JP-00118806.
XX
XX PA (CHUG-) CHUGAI RES INST MOLECULAR MEDICINE INC.
XX
XX PI Wadhwa R, Sugihara T, Yoshida A;
XX
XX DR WPI; 2000-687340/67.
XX
XX DR N-PSDB; AAC64727.
XX
XX PT Full-length tumor suppressor genes encoding Gros1-L, Gros1-S and mouse
XX PT homologs participating in regulation of cell proliferation, useful in
XX PT development of preventives and remedies of cancer.
XX
XX PS Claim 1; Page 104-107; 114pp; Japanese.
XX
XX CC The present sequence represents the mouse tumor suppressor designated
XX CC Gros1-S. Gros1-L and Gros1-S have cytostatic activity and can be used in
XX CC gene therapy. Gros1-L and Gros1-S genes are useful in the development of
XX CC drugs used to treat and prevent cancer
XX
XX SQ Sequence 542 AA;

Query Match          61.6%; Score 2384; DB 3; Length 542;
Best Local Similarity 87.9%; Pred. No. 2.2e-201;
Matches 451; Conservative 25; Mismatches 35; Indels 2; Gaps 1;

Qy 25 ESAGAGMVTPTDLLFAEGTAAYAGDWPVGVLSMERALSRAALRALRLRCRTQCAADFP 84
Db      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Qy 25 ESFPGHDVAAPDILYAGTAAYSRDRWPVGVILNMERALSRAALRALRLRCRTCAATELP 84
Db      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Qy 85 WELDPPWSPSP--AQASGAGALRDLGFFGGLLRAACLRCLGPPAAHLSSEMELEPRK 142
Db      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Qy 85 WAPDLDLGPPDPSLSDQPGAAALHDLRFFGAVLRRACLRCLGPPSAHLISELDLEFNK 144
Db      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Qy 143 RSPVNYLOVAYFKINKLEKAVAAAHFFVGNPEHMEVQONLDYVQTMVGKKEADFKDLET 202
Db      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Qy 145 RSPNYLQVAYFKINKLEKAVAAAHFFVGNPEHMEVQONLDYVQTMVGKKEADFKDLEA 204
Db      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Qy 203 QPHMQEFLGLVRLYSEEPQEAAPHLEAALQEFVAYEECRALCEGPDYDGYNYLEYNA 262
Db      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Qy 205 KPHMHEFLGLVRLYSEEPQEAAPHLEAALQEFVAYEECRALCEGPDYDGYNYLDYSA 264
Db      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Qy 263 DLFOAITDHYIOVLNCKQNCVTTELASHPSREKPFEDPLPSHNYLQFAYYNYIGNYTOAGE 322
Db      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Qy 265 DLFOAITDHYVQVLNCKQNCVTTELASHPSREKPFEDPLPSHNYLQFAYYNYIGNYTOAIE 324
Db      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Qy 323 CAKTYLLFFNDENVNQNLAYAAMLGEETHRGTPRESAKYRQRSLLLEKELFFAYDV 382
Db      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Qy 325 CAKTYLLFFNDENVHQNLAYYAMLGEEEAASSISPRENAEEYRRNLLLEKELFFAYDI 384
Db      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Qy 383 FGIPFVDPDSWTPEVTPKQLQEKQKSERETAVRISQEIQNLNKEIETLVEEKTESLDV 442
Db      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Qy 385 FGIPFVDPDSWTPEVTPKQLQEKQKSERETAVRISQEIQNLNKEIETLVEEKTESLDV 444
Db      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Qy 443 SRLTREGGPLLLEGISLTMSKLYGQVYVMDGVTSDHECQLQRLTNVAATSGDGYRG 502
Db      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Qy 445 SRLTREGGPLLLEGISLTMSKLYGQVYVMDGVTSDHECQLQRLTNVAATSGDGYRG 504
Db      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Qy 503 QTSPTTNEKFGYGVTFKALKLQEGKVPLOSA 535
Db      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Qy 505 QTSPTTNEKFGYGVTVLKALKLQEGKVPLOSA 537
Db      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
```

Search completed: July 18, 2004, 09:40:09

**This Page Blank (uspio)**

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: July 18, 2004, 09:41:58 ; Search time 62 Seconds  
(without alignments)  
3354.112 Million cell updates/sec

Title: US-10-045-815-4

Perfect score: 736  
Sequence: 1 MAVRALKLLTLLAVAAAS.....PPEPAQESUSGSSEKPKDEL 736

Scoring table:  
CLIGO  
Gapop 60.0 , Gapext 60.0

Searched: 1586107 seqs, 282547505 residues

Word size : 0

Total number of hits satisfying chosen parameters: 1586107

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database : A Geneseq\_29Jan04.\*  
1: geneseqp1980s.\*  
2: geneseqp1990s.\*  
3: geneseqp2000s.\*  
4: geneseqp2001s.\*  
5: geneseqp2002s.\*  
6: geneseqp2003as.\*  
7: geneseqp2003bs.\*  
8: geneseqp2004s.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	736	100.0	736	3	AAB36392 Human tum
2	634	96.1	736	5	ABP69129 Human pol
3	433	58.8	736	4	AAB931215 Human pro
4	433	58.8	736	4	AAB88373 Human mem
5	433	58.8	736	5	AAU09861 Novel hum
6	407	55.3	736	4	AAU09142 Human pro
7	390	53.0	592	4	AAE04244 Human gen
8	380	51.6	806	5	ABG66709 Human nov
9	380	51.6	806	5	ABG66686 Human nov
10	360	48.9	363	3	AAB36391 Human tum
11	360	48.2	804	4	AAB94482 Human pro
12	275	37.4	703	5	AAU09860 Novel hum
13	267	36.3	337	4	AAE04246 Human gen
14	244	33.2	706	5	ABP69128 Human pol
15	219	29.8	359	4	AAE04200 Human gen
16	219	29.8	359	5	ABG64502 Human alb
17	219	29.8	408	5	ABP69127 Human pol
18	188	25.5	400	5	ABG73582 Human leu
19	179	24.3	182	6	ABU70516 Human adi
20	172	23.4	217	3	AAB42734 Human ORF
21	101	13.7	173	4	AAE04217 Human gen
22	101	13.7	173	5	ABG64501 Human alb
23	91	12.4	747	3	AAB36393 Mouse tum
24	91	12.4	747	7	ABR63160 Murine po
25	82	11.1	542	3	AAB36394 Mouse tum

26	72	9.8	81	4	AAE04245 Human gen
27	60	8.2	60	4	AAE04247 Human gen
28	56	7.6	227	4	ABG16396 Novel hum
29	13	1.8	219	4	AAU87297 Novel gen
30	13	1.8	267	4	AAU17171 Novel sig
31	13	1.8	267	4	AAU87588 Novel cen
32	13	1.8	267	7	ADB93879 Human nov
33	13	1.8	282	7	ADB82712 Human pro
34	13	1.8	527	4	AAU93094 Human pro
35	13	1.8	708	6	ADA54992 Human pro
36	13	1.8	708	7	ABR63159 Human pro
37	10	1.4	286	4	ABG16279 Novel hum
38	9	1.2	15	5	ABG73583 Human leu
39	9	1.2	266	4	AAU32739 Novel hum
40	9	1.2	364	6	ABO07268 Human ps3
41	8	1.1	10	4	AAG95546 Human com
42	42	1.1	47	4	AAW94163 Human rep
43	8	1.1	47	4	ABR95628 Human tes
44	8	1.1	222	6	ABU26511 Protein e
45	8	1.1	264	6	ABU11680 Human MDD

#### ALIGNMENTS

##### RESULT 1

AAAB36392  
ID AAB36392 standard; protein; 736 AA.  
XX  
XX AAB36392;  
AC  
XX  
DT 27-FEB-2001 (first entry)  
XX  
DE Human tumour suppressor Gros1-S protein SEQ ID NO:4.  
XX  
KW Tumour suppressor; Gros1-L; Gros1-S; cell proliferation; regulation;  
KW cancer; cytostatic; gene therapy.  
XX  
OS Homo sapiens.  
XX  
PN WO200065047-A1.  
XX  
PD 02-NOV-2000.  
XX  
PF 26-APR-2000; 2000WO-JP002731.  
XX  
PR 26-APR-1999; 95JP-00118806.  
XX  
(CHUG-) CHUGAI RES INST MOLECULAR MEDICINE INC.  
XX  
PI Wadhwa R, Sugihara T, Yoshida A;  
XX  
DR WPI; 2000-687340/67.  
DR N-PSDB; AAC64725.  
XX  
PT Full-length tumor suppressor genes encoding Gros1-L, Gros1-S and mouse  
PT homologs participating in regulation of cell proliferation, useful in  
PT development of preventives and remedies of cancer.  
XX  
PS Claim 1; Page 78-83; 114pp; Japanese.  
XX  
CC The present sequence represents the human tumour suppressor designated  
CC Gros1-S. Gros1-L and Gros1-S have cytostatic activity and can be used in  
CC Gene therapy. Gros1-L and Gros1-S genes are useful in the development of  
CC drugs used to treat and prevent cancer  
XX  
SQ Sequence 736 AA;

Query Match 100.0%; Score 736; DB 3; Length 736;  
Best Local Similarity 100.0%; Pred. No. 0;  
Matches 736; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1 MAVRALKLLTLLAVAAASQAQVESEAGWGVTPDLLFAEGTAAYAGDWPGVLSMER 60

Db	1	MAVRALKLLTTLAAVAAASQAEVSEBAGMVTPTDLLFAEGTAAYARGDWPQVLSMER	60
QY	61	ALRSRAALRALRLCRCTCCAADPWLDPWSPPAQSCAGALRDLSPFGGLLRAACL	120
Db	61	ALRSRAALRALRLCRCTCCAADPWLDPWSPPAQSCAGALRDLSPFGGLLRAACL	120
QY	121	RRCLGPPAAHSLSEMELEFRKSPYNYLOVAFKINKLEKAVAAAHTFFVGNPEMEMQ	180
Db	121	RRCLGPPAAHSLSEMELEFRKSPYNYLOVAFKINKLEKAVAAAHTFFVGNPEMEMQ	180
QY	181	QNLDDYQTMGSKVKEADPKDLSTQPHMQEFLGVLRLYSEEQPOEAVPHLEALQYFYAYE	240
Db	181	QNLDDYQTMGSKVKEADPKDLSTQPHMQEFLGVLRLYSEEQPOEAVPHLEALQYFYAYE	240
QY	241	ECRALCEGPDYDGYNYLYNADLFOAITDHYIQVLNCKQNCVTELASHPSREKPPDFL	300
Db	241	ECRALCEGPDYDGYNYLYNADLFOAITDHYIQVLNCKQNCVTELASHPSREKPPDFL	300
QY	301	PSHNYLQFAYNYTNGYTQAGECAKTYLLFPFNDVNNQNLAYYAAMLGSEHTRSIGPRE	360
Db	301	PSHNYLQFAYNYTNGYTQAGECAKTYLLFPFNDVNNQNLAYYAAMLGSEHTRSIGPRE	360
QY	361	SAKEYRORSLEKELLFPAYDFGIPFVDPSWTPPEEVIKRLQEKOKSERETAVRISQE	420
Db	361	SAKEYRORSLEKELLFPAYDFGIPFVDPSWTPPEEVIKRLQEKOKSERETAVRISQE	420
QY	421	IGNLMKEIETLVEBKTESLDVSLTRREGGPLYEGISLTWNKLNQYQVVMQGVISD	480
Db	421	IGNLMKEIETLVEBKTESLDVSLTRREGGPLYEGISLTWNKLNQYQVVMQGVISD	480
QY	481	HECELOQLTNVAATSGDVGRCQTSPTPNEKFGYVTVFKALKGOEKGKVPLOSAHLYN	540
Db	481	HECELOQLTNVAATSGDVGRCQTSPTPNEKFGYVTVFKALKGOEKGKVPLOSAHLYN	540
QY	541	VTEKVRIMESYFRDLTPLYFSYSHLVCRTAIEBQVQAEKDDSHPHVDNCILNAETLVC	600
Db	541	VTEKVRIMESYFRDLTPLYFSYSHLVCRTAIEBQVQAEKDDSHPHVDNCILNAETLVC	600
QY	601	VKEPPAYTFRYSAILYLINGDFDGNFYFTELDAKTYTAEVQPCGKRAVGFSGTENPHG	660
Db	601	VKEPPAYTFRYSAILYLINGDFDGNFYFTELDAKTYTAEVQPCGKRAVGFSGTENPHG	660
QY	661	VKATRGORCAIALWFTLDRHSERDRVQADDLVKMLFSPSEMDLSOEQLDAQOGPPEP	720
Db	661	VKATRGORCAIALWFTLDRHSERDRVQADDLVKMLFSPSEMDLSOEQLDAQOGPPEP	720
QY	721	AOESLSGSESKPKDEL 736	
Db	721	AOESLSGSESKPKDEL 736	
RESULT 2			
ID	ABP69129	standard; protein; 736 AA.	
XX	AC	ABP69129;	
XX	AC	ABP69129;	
DT	20-JAN-2003	(first entry)	
DE	Human	polypeptide SEQ ID NO 1176.	
KW	Human;	genome mapping; gene therapy; food supplement; virus; fungus;	
KW	cell-proliferative disorder; neurodegenerative disease; bacterial;		
KW	Parkinson's disease; Alzheimer's disease; autoimmune disease;		
KW	multiple sclerosis; diabetes; genetic disorder; wound; burn; infection;		
KW	arthritis; cytostatic; immunomodulator; nootropic; neuroprotective;		
KW	antiParkinsonian; antidiabetic; immunosuppressive; dermatological;		
KW	haemostatic; vulnerary; fungicide; antibacterial; virucide; protozoacide;		
OS	Homo sapiens.		
XX			

PN	WC200270539-A2.		
XX	12-SEP-2002.		
XX	05-MAR-2002; 2002WO-US0005095.		
XX	05-MAR-2002; 2001US-00799451.		
XX	(HYSE-) HYSEQ INC.		
PI	Tang YT, Zhou P, Goodrich RW, Asundi V, Zhang J, Zhao QA, Ren F;		
PI	Xue AJ, Yang Y, Ma Y, Yamazaki V, Chen R, Wang Z, Ghosh M;		
PI	Wehrman T, Wang J, Wang D, Drmanac R;		
XX	WPI; 2002-759812/82.		
DR	N-PSDB; ABZ11346.		
XX	New polynucleotides comprising sequences assembled from expressed		
PT	sequence tags (ESTs), useful for treating cell-proliferative,		
PT	neurodegenerative, autoimmune, genetic, myeloid or lymphoid, or platelet		
PT	or coagulation disorders.		
XX	Claim 9; SEQ ID NO 1176; 1012bp + Sequence Listing; English.		
PS	The invention relates to an isolated polynucleotide (I) comprising a		
CC	nucleotide sequence selected from any of 948 sequences (ABZ11119-		
CC	ABZ12066) or their mature protein coding portion, active domain coding		
CC	protein or complementary sequences. The polynucleotides are useful for		
CC	identifying expressed genes or for physical mapping of human genome. The		
CC	encoded polypeptides (ABP6902-ABP6949) are useful as molecular weight		
CC	markers, as a food supplement, for generating antibodies, in medical		
CC	imaging, screening and diagnostic assays and for treating cell-		
CC	proliferative disorders (cancer), neurodegenerative diseases (Parkinson's		
CC	or Alzheimer's disease), autoimmune diseases (multiple sclerosis,		
CC	diabetes, lupus) genetic disorders, myeloid or lymphoid disorders,		
CC	platelet or coagulation disorders, wound, burns, incision, ulcers, liver		
CC	or lung fibrosis, infections (bacterial, viral, fungal, parasitic),		
CC	arthritis, etc. Note: The sequence data for this patent did not form part		
CC	of the printed specification, but was obtained in electronic format		
CC	directly from WIPO at fip.wipo.int/pub/published_pct_sequences		
XX	Sequence 736 AA;		
SQ			
Query Match 86.1%; Score 634; DB 5; Length 736;			
Best Local Similarity 100.0%; Pred. No. 0;			
Matches 634; Conservative 0; Mismatches 0; Indels 0; Gaps 0;			
QY	103	ALRDLSPFGGLLRAACLRCLGPPAAHSLSEMELEFRKSPYNYLOVAFKINKLEKA	162
Db	103	ALRDLSPFGGLLRAACLRCLGPPAAHSLSEMELEFRKSPYNYLOVAFKINKLEKA	162
QY	163	VAAAHTEFFVGNPEHEMOMQNDYYQTMGSKVKEADPKDLSTQPHMQEFLGVLRLYSEEQPQ	222
Db	163	VAAAHTEFFVGNPEHEMOMQNDYYQTMGSKVKEADPKDLSTQPHMQEFLGVLRLYSEEQPQ	222
QY	223	EAVPHLEALQYFYVAYEBCRALCEGPDYDGYNYLYNADLFOAITDHYIQVLNCKQNC	282
Db	223	EAVPHLEALQYFYVAYEBCRALCEGPDYDGYNYLYNADLFOAITDHYIQVLNCKQNC	282
QY	283	VTELASHPSREKPPDFLPSHNYLQFAYNYTNGYTQAGECAKTYLLFPFNDVNNQNL	342
Db	283	VTELASHPSREKPPDFLPSHNYLQFAYNYTNGYTQAGECAKTYLLFPFNDVNNQNL	342
QY	343	YYAAMLGSEHTRSIGPRESAKYRORSLEKELLFPAYDFGIPVDPSWTPPEEVIKPR	402
Db	343	YYAAMLGSEHTRSIGPRESAKYRORSLEKELLFPAYDFGIPVDPSWTPPEEVIKPR	402
QY	403	LOEKOKSERETAVRISQIEGNLMKEIETLVEBKTESLDVSLTRREGGPLYEGISLTWN	462
Db	403	LOEKOKSERETAVRISQIEGNLMKEIETLVEBKTESLDVSLTRREGGPLYEGISLTWN	462
QY	463	SKLLNGYQVVMQGVISDHECOELQRLTNVAATSGDVGRCQTSPTPNEKFGYVTVFKAL	522



Db 463 SKLLNGYQVRVMDGVISDHECQELQRLTNVAATSGDYGRTSPHTPNEKFGVTVFKAL 522

QY 523 KLGOEGKVPLOSALHYNNVTEKVRIMESYFRDLTPLYFSYSHLVCRTAIEEVOAERKDD 582

Db 523 KLGOEGKVPLOSALHYNNVTEKVRIMESYFRDLTPLYFSYSHLVCRTAIEEVOAERKDD 582

QY 583 SHPVHVDNCLINAEITLVCVKEPPAYTFRDYSALLYLNGDFDGNFFFTTDLDAKTVAEQ 642

Db 583 SHPVHVDNCLINAEITLVCVKEPPAYTFRDYSALLYLNGDFDGNFFFTTDLDAKTVAEQ 642

QY 643 PQCGRAVGSSTENPHGVKAVTRGORCAIALWFTLDRHSDRDRVQADLVKMLFSPPE 702

Db 643 PQCGRAVGSSTENPHGVKAVTRGORCAIALWFTLDRHSDRDRVQADLVKMLFSPPE 702

QY 703 MDLSQQLDQAQGPPEPAQESLSGSSEKPKDEL 736

Db 703 MDLSQQLDQAQGPPEPAQESLSGSSEKPKDEL 736

RESULT 3

AAB93215

ID AAB93215 standard; protein; 736 AA.

XX AAB93215;

AC AAB93215;

XX 26-JUN-2001 (first entry)

XX Human protein sequence SEQ ID NO:12194.

XX Human; primer; detection; diagnosis; antisense therapy; gene therapy.

XX Homo sapiens.

OS EP1074617-A2.

XX PD 07-FEB-2001.

XX PF 28-JUL-2000; 2000EP-00116126.

XX PR 29-JUL-1999; 99JP-00248036.

XX PR 27-AUG-1999; 99JP-00300253.

PR 11-JAN-2000; 2000JP-00118776.

PR 02-MAY-2000; 2000JP-00183767.

PR 09-JUN-2000; 2000JP-00241899.

XX (HELI-) HELIX RES INST.

XX Ota T, Isogai T, Nishikawa T, Hayashi K, Saito K, Yamamoto J;

PI Ishii S, Sugiyama T, Wakamatsu A, Nagai K, Otsuki T;

XX WPI; 2001-318749/34.

XX Primer sets for synthesizing polynucleotides, particularly the 5602 full-length cDNAs defined in the specification, and for the detection and/or diagnosis of the abnormality of the proteins encoded by the full-length cDNAs.

XX Claim 8; SEQ ID NO 12194; 2537pp + Sequence Listing; English.

XX The present invention describes primer sets for synthesizing 5602 full-length cDNAs defined in the specification. Where a primer set comprises: (a) an oligo-dT primer and an oligonucleotide complementary to the complementary strand of a polynucleotide which comprises one of the 5602 nucleotide sequences defined in the specification, where the oligonucleotide comprises at least 15 nucleotides; or (b) a combination of an oligonucleotide comprising a sequence complementary to the complementary strand of a polynucleotide which comprises a 5'-end sequence and an oligonucleotide comprising a sequence complementary to a polynucleotide which comprises a 3'-end sequence, where the oligonucleotide comprises at least 15 nucleotides and the combination of the 5'-end sequence/3'-end sequence is selected from those defined in the specification. The primer sets can be used in antisense therapy and in gene therapy. The primers are useful for synthesizing polynucleotides,

CC particularly full-length cDNAs. The primers are also useful for the detection and/or diagnosis of the abnormality of the proteins encoded by the full-length cDNAs. The primers allow obtaining of the full-length cDNAs easily without any specialised methods. AAB93166 to AAB9328 and AAB9329 to AAB9342 represent human cDNA sequences; AAB92446 to AAB95893 represent human amino acid sequences; and AAB93629 to AAB93632 represent oligonucleotides, all of which are used in the exemplification of the present invention

XX Sequence 736 AA;

SQ

Query Match 58.8%; Score 433; DB 4; Length 736;

Best Local Similarity 99.6%; Pred. No. 0;

Matches 733; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 MAVRALKLLTLLAVAAASQAEVESEAGWGMVTPDLLFAEGTAAYARGDMPGVVLSNER 60

Db 1 MAVRALKLLTLLAVAAASQAEVESEAGWGMVTPDLLFAEGTAAYARGDMPGVVLSNER 60

QY 61 ALRSRAALRALRLRCRTQCAADFPWELDPDWSPPAQASGAGALRLDSFFGGLLRAACL 120

Db 61 ALRSRAALRALRLRCRTQCAADFPWELDPDWSPPAQASGAGALRLDSFFGGLLRAACL 120

QY 121 RRCIGPPAAHSISEMELEFRKSPYNYLVAYFKINKLEKAVAAAHFFVGNDEHMQ 180

Db 121 RRCIGPPAAHSISEMELEFRKSPYNYLVAYFKINKLEKAVAAAHFFVGNDEHMQ 180

QY 181 QNLDYYQTMGSKVKEADFKDLETPHMQEFLGVLYSEEQPQEAHPHLEALQYFYVAYE 240

Db 181 QNLDYYQTMGSKVKEADFKDLETPHMQEFLGVLYSEEQPQEAHPHLEALQYFYVAYE 240

QY 241 ECRALCEGPDYDGYNYLEYNADLFQAITDHYIQLNCKQNCVTELASHPSREKPFDFL 300

Db 241 ECRALCEGPDYDGYNYLEYNADLFQAITDHYIQLNCKQNCVTELASHPSREKPFDFL 300

QY 301 PSHYNYLOFAYNYIGNYTQAGECAKTYLLFFPNDEVMNQLAYYAAMLGEHTRSIGPRE 360

Db 301 PSHYNYLOFAYNYIGNYTQAGECAKTYLLFFPNDEVMNQLAYYAAMLGEHTRSIGPRE 360

QY 361 SAKYRQSRSLLEKELLFPAYDVFGIPFVDPDSWTPPEVPIKRLQEKQKSERETAVRISQE 420

Db 361 SAKYRQSRSLLEKELLFPAYDVFGIPFVDPDSWTPPEVPIKRLQEKQKSERETAVRISQE 420

QY 421 IGMLMKEITLVEKTKESLDVSRLTREGGPLYEGISLTWNKLLANGYQVRVMDGVISD 480

Db 421 IGMLMKEITLVEKTKESLDVSRLTREGGPLYEGISLTWNKLLANGYQVRVMDGVISD 480

QY 481 HECQELQRLTNVAATSGDYGRTSPHTPNEKFGVTVFKALQGOEGKVPLOSALHYNN 540

Db 481 HECQELQRLTNVAATSGDYGRTSPHTPNEKFGVTVFKALQGOEGKVPLOSALHYNN 540

QY 541 VTEKVRIMESYFRDLTPLYFSYSHLVCRTAIEEVOAERKDDSHPVHVDNCLINAEITLVC 600

Db 541 VTEKVRIMESYFRDLTPLYFSYSHLVCRTAIEEVOAERKDDSHPVHVDNCLINAEITLVC 600

QY 601 VKEPPAYTFRDYSALLYLNGDFDGNFFFTTDLDAKTVAEQPOCGRAVGSSTENPHG 660

Db 601 VKEPPAYTFRDYSALLYLNGDFDGNFFFTTDLDAKTVAEQPOCGRAVGSSTENPHG 660

QY 661 VKAVTRGORCAIALWFTLDRHSDRDRVQADLVKMLFSPPEMDLSQQLDQAQGPPEP 720

Db 661 VKAVTRGORCAIALWFTLDRHSDRDRVQADLVKMLFSPPEMDLSQQLDQAQGPPEP 720

QY 721 AQESLSGSSEKPKDEL 736

Db 721 AQESLSGSSEKPKDEL 736

RESULT 4

AAB88373

ID AAB88373 standard; protein; 736 AA.

XX AAB88373;

AC AAB88373;

XX DT 23-MAY-2001 (first entry)  
 XX DE Human membrane or secretory protein clone PSEC0109.  
 XX KW Human; secretory protein; membrane protein; vaccine; gene therapy;  
 XX KW rheumatoid arthritis; diabetes.  
 XX OS Homo sapiens.  
 XX FN EP1067182-A2.  
 XX PD 10-JAN-2001.  
 XX PF 07-JUL-2000; 2000EP-00114090.  
 XX PR 08-JUL-1999; 99JP-00194179.  
 XX PR 11-JAN-2000; 2000JP-00118775.  
 XX PR 02-MAY-2000; 2000JP-00183766.  
 XX PA (HELI-) HELIX RES INST.  
 XX Ota T, Isogai T, Nishikawa T, Kawai Y, Sugiyama T, Hayashi K;  
 WPI; 2001-093989/11.  
 XX N-PSDB; AAF93800.  
 XX Nucleic acids encoding secretory proteins/membrane proteins, useful in  
 PT gene therapy or as candidate target molecules in drug development.  
 XX Claim 1; SEQ ID NO 114; 609pp + Sequence Listing; English.  
 XX This invention relates to nucleic acid sequences AAF93744 - AAF93916  
 CC which encode human secretory or membrane proteins represented by AAB88317  
 CC - AAB88419. Included in the invention are primers AAF93917 - AAF94295 and  
 CC AAF62232 - AAF62235 which are used to isolate the cDNA sequences of the  
 CC invention. The invention also includes methods for the production of  
 CC antibodies directed against the proteins, and cDNA sequences, which can  
 CC be used in vaccines. The polynucleotide sequences can be used in gene  
 CC therapy. The polynucleotide sequences and the proteins they encode may be  
 CC used in the prevention, treatment and diagnosis of diseases associated  
 CC with inappropriate secretory protein/membrane protein expression. The  
 CC nucleic acids and complementary sequences may also be used as DNA probes  
 CC in diagnostic assays (e.g. polymerase chain reactions (PCR)) to detect  
 CC and quantitate the presence of similar nucleic acid sequences in samples.  
 CC They may also be used to study the expression and function of secretory  
 CC proteins/membrane polypeptides and their role in metabolism. The  
 CC polypeptides may be used as antigens in the production of antibodies  
 CC against them and in assays to identify modulators (agonists and  
 CC antagonists) of expression and activity. The antibodies and antagonists  
 CC may also be used as therapeutic agents to down regulate expression and  
 CC activity. The antibodies may also be used as diagnostic agents for  
 CC detecting the presence of the polypeptides in samples (e.g. by enzyme  
 CC linked immunosorbant assay (ELISA)). Examples of diseases which may be  
 CC treated include rheumatoid arthritis and diabetes  
 XX Sequence 736 AA;  
 SQ Query Match 58.8%; Score 433; DB 4; Length 736;  
 Best Local Similarity 99.6%; Pred. No. 0;  
 Matches 733; Conservative 0; Mismatches 3; Indels 0; Gaps 0;  
 Qy 1 MAVFALKLLTLLAVAAQAQVESEAGMGWTPDLLFAEGTAAYARGDWPQVLSMER 60  
 Db 1 MAVFALKLLTLLAVAAQAQVESEAGMGWTPDLLFAEGTAAYARGDWPQVLSMER 60  
 Qy 61 ALRSEALRALRLRCRTQCAADFPWELDPDWSFSPAQAGALRDLSPFGGLLRAACL 120  
 Db 61 ALRSEALRALRLRCRTQCAADFPWELDPDWSFSPAQAGALRDLSPFGGLLRAACL 120  
 Qy 121 RRCILGPPAAHLSSEMELEFRKSPYNYLQVAFKINKLEKAVAAAHFFVGNPEHMEQ 180  
 Db 121 RRCILGPPAAHLSSEMELEFRKSPYNYLQVAFKINKLEKAVAAAHFFVGNPEHMEQ 180

QY 181 QNLDYYQTMGSKVEADFKDLETPHMQEFLRGVRLYSEEQPQEAHPHLEALQEFVAYE 240  
 DB 181 QNLDYYQTMGSKVEADFKDLETPHMQEFLRGVRLYSEEQPQEAHPHLEALQEFVAYE 240  
 QY 241 ECRALCEGPDYDGYNYLEYNADLFOAITDHYIQLVNLCKQNCVTELASHSPREKPFDFL 300  
 DB 241 ECRALCEGPDYDGYNYLEYNADLFOAITDHYIQLVNLCKQNCVTELASHSPREKPFDFL 300  
 QY 301 PSHYNYLOFAYYNYGNITQAGECAKTYLLFPFNDVNMNQLAYYAAMLGBEHTSISGPRE 360  
 DB 301 PSHYNYLOFAYYNYGNITQAGECAKTYLLFPFNDVNMNQLAYYAAMLGBEHTSISGPRE 360  
 QY 361 SAKYRQBSLLEKELLFPAYDVFGIPFVDPDSWTPEEVIKRLQEKOKSERETAVRISOE 420  
 DB 361 SAKYRQBSLLEKELLFPAYDVFGIPFVDPDSWTPEEVIKRLQEKOKSERETAVRISOE 420  
 QY 421 IGNLMKEITLVEEKTESLDVSRLTREGGPLLYEGISLTWNKSLNGYQRVNMDGVISD 480  
 DB 421 IGNLMKEITLVEEKTESLDVSRLTREGGPLLYEGISLTWNKSLNGYQRVNMDGVISD 480  
 QY 481 HECQELQRLTNVAATSGDGYRGQTSPTPNKFKYGVTVFKALKLGQSKVPLQSAHLYYN 540  
 DB 481 HECQELQRLTNVAATSGDGYRGQTSPTPNKFKYGVTVFKALKLGQSKVPLQSAHLYYN 540  
 QY 541 VTEKVRIMESYFRDLTPLYFSYSHLVCRCTAIEEVQAEKDDSHPVHVDNCLNAETLVC 600  
 DB 541 VTEKVRIMESYFRDLTPLYFSYSHLVCRCTAIEEVQAEKDDSHPVHVDNCLNAETLVC 600  
 QY 601 VKEPAYTFRDYSAILYLNGDFDGNFYFTLDAKTVAETAEVQPCGAVGSSSTENPHG 660  
 DB 601 VKEPAYTFRDYSAILYLNGDFDGNFYFTLDAKTVAETAEVQPCGAVGSSSTENPHG 660  
 QY 661 VKAVTRGQRCALALWFTLDRHSERDRVQADDLVKMLFSPPEMDLSOEQPLDAQGGPPEP 720  
 DB 661 VKAVTRGQRCALALWFTLDRHSERDRVQADDLVKMLFSPPEMDLSOEQPLDAQGGPPEP 720  
 QY 721 AQESLSGSESKPKDEL 736  
 DB 721 AQESLSGSESKPKDEL 736  
 RESULT 5  
 AAU09861  
 ID AAU09861 standard; protein; 736 AA.  
 XX  
 AC AAU09861;  
 XX  
 DT 26-FEB-2002 (first entry)  
 XX  
 DE Novel human secreted protein #2.  
 XX  
 KW Secreted protein; cytostatic; immunosuppressive; vulnary; vaccine;  
 KW antinflammatory; neuroprotective; nephrotropic; cardiovascular; human;  
 KW cancer; autoimmune disease; wound healing disorder; infection;  
 KW haematopoietic disorder; inflammatory disorder; infertility;  
 KW neurological disease; psychiatric disease; cardiovascular disease;  
 KW respiratory disease; renal; gastrointestinal.  
 XX  
 OS Homo sapiens.  
 XX  
 XX WO200179454-A1.  
 XX  
 XX 25-OCT-2001.  
 XX  
 PF 11-APR-2001; 2001WO-US011797.  
 XX  
 PR 13-APR-2000; 2000US-0196603P.  
 PR 24-APR-2000; 2000US-0199417P.  
 XX  
 XX (SMIK ) SMITHKLINE BEECHAM CORP.  
 PA (SMIK ) SMITHKLINE BEECHAM PLC.  
 XX

PI Agarwal P, Murdoch PR, Rizvi SK, Smith RF, Xiang Z;  
XX WPI; 2002-061975/08.  
DR N-PSDB; AAS17573.  
XX  
PT New secreted proteins or polypeptides, useful for treating e.g. cancer,  
PT autoimmune diseases, wound healing disorder, infections, hematopoietic  
PT disorders, inflammatory disorders, infertility, cancer.  
XX  
PS Claim 1; Page 54-56; 92pp; English.

XX The invention relates to an isolated novel secreted polypeptide (I) and  
XX polynucleotide (II). (I) and (II) are useful for treating cancer,  
CC autoimmune diseases, wound healing disorder, infections, hematopoietic  
CC disorders, inflammatory disorders, infertility, neurological and  
CC psychiatric diseases, cardiovascular diseases, respiratory diseases,  
CC renal diseases, or gastrointestinal diseases. These may also be used to  
CC treat diseases, abnormalities and disorders caused by abnormal  
CC expression, production, function and/or metabolism of the genes, as  
CC vaccines for inducing immunological response in a mammal, and in  
CC screening methods for detecting the effect of added compounds on the  
CC production of mRNA and polypeptide in cells. The polypeptides can be used  
CC as immunogens to produce antibodies immunospecific for the polypeptides,  
CC and to identify membrane-bound or soluble receptors. The polynucleotides  
CC may be used as diagnostic reagents, in chromosome localisation studies,  
CC and in tissue expression studies. The present sequence represents the  
CC amino acid sequence of novel human secreted protein #2  
XX  
SQ Sequence 736 AA;

Query Match 58.8%; Score 433; DB 5; Length 736;  
Best Local Similarity 99.6%; Pred. No. 0;  
Matches 733; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 MAVRALKLLTLLAVAAASQAQVESEAGAGMTPTDLLFAEGTAAYARGDWPVGLSMER 60  
DB 1 MAVRALKLLTLLAVAAASQAQVESEAGAGMTPTDLLFAEGTAAYARGDWPVGLSMER 60  
QY 61 ALRSRAALRALRLCRTOCAADFPWELDPWSPSPAQAGALRDLSPFGGLLRAACL 120  
DB 61 ALRSRAALRALRLCRTOCAADFPWELDPWSPSPAQAGALRDLSPFGGLLRAACL 120  
QY 121 RRCGLGPAASLSSEMELEFRKSPYNYLOVAFKINKLEKAVAAHTFFVGNPEHMEQ 180  
DB 121 RRCGLGPAASLSSEMELEFRKSPYNYLOVAFKINKLEKAVAAHTFFVGNPEHMEQ 180  
QY 181 QNLDDYQTMGSKVEADFKDLETPHMQEFLGLVRLYSEEQPQAVPHLEALQEFYVAYE 240  
DB 181 QNLDDYQTMGSKVEADFKDLETPHMQEFLGLVRLYSEEQPQAVPHLEALQEFYVAYE 240  
QY 241 ECRALCGPYDYGNYLEYNADLFQAITDHYIQLVNLCKQNCVTELASHPSREKPFEDFL 300  
DB 241 ECRALCGPYDYGNYLEYNADLFQAITDHYIQLVNLCKQNCVTELASHPSREKPFEDFL 300  
QY 301 PSHVNYLOFAYNIGNYTOAGECAKTYLLFPNDVNMQLAYYAAMLGEHTRSIGPRE 360  
DB 301 PSHVNYLOFAYNIGNYTOAGECAKTYLLFPNDVNMQLAYYAAMLGEHTRSIGPRE 360  
QY 361 SAKERYORSLLLEKLLFFAYDVGIPVDPSMTPEEVIKRLQEKQKSRRETAVRISQE 420  
DB 361 SAKERYORSLLLEKLLFFAYDVGIPVDPSMTPEEVIKRLQEKQKSRRETAVRISQE 420  
QY 421 IGNLMKEIETLVEBKTESLDVSLTREGGLYEGISLTWNSKLLNGYRVMDGVISD 480  
DB 421 IGNLMKEIETLVEBKTESLDVSLTREGGLYEGISLTWNSKLLNGYRVMDGVISD 480  
QY 481 HECQELQRLTNVAATSGDVGRTGQTSPTPNEKFYGVTVFFKALIKLQEGKVPLOSAHLYN 540  
DB 481 HECQELQRLTNVAATSGDVGRTGQTSPTPNEKFYGVTVFFKALIKLQEGKVPLOSAHLYN 540  
QY 541 VTEKVRIMESYFRLLDPLFYSHLYCRTAIEVQAEKDDSHPHVNDNCILNAETLYC 600  
DB 541 VTEKVRIMESYFRLLDPLFYSHLYCRTAIEVQAEKDDSHPHVNDNCILNAETLYC 600

QY 601 VKEPPAYTRDYSAIYLNGDFDGGNFYFTELDAKTVTAEVQPCGAVGSSGTENPHG 660  
DB 601 VKEPPAYTRDYSAIYLNGDFDGGNFYFTELDAKTVTAEVQPCGAVGSSGTENPHG 660  
QY 661 VKAVTRGRCALALWFTLDPHRSERDRVQADLVXVLFSPPEMDLSQOPLDAQGPPEP 720  
DB 661 VKAVTRGRCALALWFTLDPHRSERDRVQADLVXVLFSPPEMDLSQOPLDAQGPPEP 720  
QY 721 AQESLSGSESKEPKDEL 736  
DB 721 AQESLSGSESKEPKDEL 736

# RESULT 6

AAB93142  
ID AAB93142 standard; protein; 736 AA.

AC AAB93142;

DT 26-JUN-2001 (first entry)

DE Human protein sequence SEQ ID NO:12045.

XX Human; primer; detection; diagnosis; antisense therapy; gene therapy.

XX Homo sapiens.

PN EP1074617-A2.

PD 07-FEB-2001.

PF 28-JUL-2000; 2000EP-00116126.

PR 29-JUL-1999; 99JP-00248036.

PR 27-AUG-1999; 99JP-00300253.

PR 11-JAN-2000; 2000JP-00118776.

PR 02-MAY-2000; 2000JP-00183767.

PR 09-JUN-2000; 2000JP-00241899.

XX (HELI-) HELIX RES INST.

PI Ota T, Isogai T, Nishikawa T, Hayashi K, Saito K, Yamamoto J;

PI Ishii S, Sugiyama T, Wakamatsu A, Nagai K, Otsuki T;

XX WPI; 2001-318749/34.

PT Primer sets for synthesizing polynucleotides, particularly the 5602 full-length cDNAs defined in the specification, and for the detection and/or diagnosis of the abnormality of the proteins encoded by the full-length cDNAs.

PS Claim 8; SEQ ID NO 12045; 2537pp + Sequence Listing; English.

XX The present invention describes primer sets for synthesizing 5602 full-length cDNAs defined in the specification. Where a primer set comprises: (a) an oligo-dT primer and an oligonucleotide complementary to the complementary strand of a polynucleotide which comprises one of the 5602 nucleotide sequences defined in the specification, where the oligonucleotide comprises at least 15 nucleotides; or (b) a combination of an oligonucleotide comprising a sequence complementary to the complementary strand of a polynucleotide which comprises a 5'-end sequence and an oligonucleotide comprising a sequence complementary to a polynucleotide which comprises a 3'-end sequence, where the oligonucleotide comprises at least 15 nucleotides, and the combination of the 5'-end sequence/3'-end sequence is selected from those defined in the specification. The primer sets can be used in antisense therapy and in gene therapy. The primers are useful for synthesizing polynucleotides, particularly full-length cDNAs. The primers are also useful for the detection and/or diagnosis of the abnormality of the proteins encoded by the full-length cDNAs. The primers allow obtaining of the full-length cDNAs easily without any specialised methods. AAH03166 to AAH13628 and AAH13633 to AAH18742 represent human cDNA sequences; AAB92446 to AAB95893

CC represent human amino acid sequences; and AAHL3629 to AAHL3632 represent  
 CC oligonucleotides, all of which are used in the exemplification of the  
 CC present invention  
 XX Sequence 736 AA;  
 SQ

Query Match 55.3%; Score 407; DB 4; Length 736;  
 Best Local Similarity 99.6%; Pred. No. 0;  
 Matches 707; Conservative 0; Mismatches 3; Indels 0; Gaps 0;  
 QY 1 NAVRALKLLTLLAVVAAASQAEESEAGWGVTPDLLFABGTAAYARGDWPVGLSMER 60  
 DB 1 NAVRALKLLTLLAVVAAASQAEESEAGWGVTPDLLFABGTAAYARGDWPVGLSMER 60  
 QY 61 ALRGAALRALRLCRTOCAADFPEWLDPDWSPQAASGAGALRDLSPFGGLLRAACL 120  
 DB 61 ALRGAALRALRLCRTOCAADFPEWLDPDWSPQAASGAGALRDLSPFGGLLRAACL 120  
 QY 121 RRCGLPPAAHLSBEMEELFRKSPYNYLQVAYFKINKLEKAVAAAHFTFFVGNPEHMEMQ 180  
 DB 121 RRCGLPPAAHLSBEMEELFRKSPYNYLQVAYFKINKLEKAVAAAHFTFFVGNPEHMEMQ 180  
 QY 181 QNLDYQTMGSKGKADFQDLTQPHMQEFLRGVRLYSEEQPQEAHPHLEALQEFYFAYE 240  
 DB 181 QNLDYQTMGSKGKADFQDLTQPHMQEFLRGVRLYSEEQPQEAHPHLEALQEFYFAYE 240  
 QY 241 ECRALCEGPDYDGYNYLEYNADLFOAITDHYIQVLNCKQNCVTELASHPSREKPFDFL 300  
 DB 241 ECRALCEGPDYDGYNYLEYNADLFOAITDHYIQVLNCKQNCVTELASHPSREKPFDFL 300  
 QY 301 PSHNYLQFAYNYNGYTOGECATYLLFPNDENVNQNLAAYAAMLGBEHTRSIGPRE 360  
 DB 301 PSHNYLQFAYNYNGYTOGECATYLLFPNDENVNQNLAAYAAMLGBEHTRSIGPRE 360  
 QY 361 SAKVQRSLSEKELLFPAYDVGIFPVDPSWTPPEVVKRLOEKOKSERETAVRISQE 420  
 DB 361 SAKVQRSLSEKELLFPAYDVGIFPVDPSWTPPEVVKRLOEKOKSERETAVRISQE 420  
 QY 421 IGNLMKEITLVEBKTESLDVSLRTREGGPLYEGISLTMNSKLLNGYQRVMDGVISD 480  
 DB 421 IGNLMKEITLVEBKTESLDVSLRTREGGPLYEGISLTMNSKLLNGYQRVMDGVISD 480  
 QY 481 HECCLOQLTNVAATSGDYGEGTSPHTPNEKFGVTVFKALKLGQSGKVPLOSALHYLN 540  
 DB 481 HECCLOQLTNVAATSGDYGEGTSPHTPNEKFGVTVFKALKLGQSGKVPLOSALHYLN 540  
 QY 541 VTEKVRIMESYFRILDTPLYSYSHLVCRTAIEBQVQERKDDSHPVHVDNCILNAETLVC 600  
 DB 541 VTEKVRIMESYFRILDTPLYSYSHLVCRTAIEBQVQERKDDSHPVHVDNCILNAETLVC 600  
 QY 601 VKEPPAYTFRDYSAILYLNGDFDGNFFYFELDAKTTVAEQPQCGRAVGFSSGTENPHG 660  
 DB 601 VKEPPAYTFRDYSAILYLNGDFDGNFFYFELDAKTTVAEQPQCGRAVGFSSGTENPHG 660  
 QY 661 VKAVTRGORCAIALWFLTPHRSERDRVQADDLVKMLFSPPEMDLSQEQP 710  
 DB 661 VKAVTRGORCAIALWFLTPHRSERDRVQADDLVKMLFSPPEMDLSQEQP 710

RESULT 7  
 AA04244  
 ID AA04244 standard; protein; 592 AA.  
 XX  
 AC AA04244;  
 XX  
 DT 09-AUG-2001 (first entry)  
 XX  
 DE Human gene 2 encoded secreted protein fragment, SEQ ID NO:100.  
 XX  
 KW Human; secreted protein; proliferative disorder; cancer; tumour;  
 KW fetal abnormality; developmental abnormality; haematopoietic disorder;  
 KW immune system disorder; AIDS; autoimmune disease; rheumatoid arthritis;  
 KW inflammation; allergy; neurological disorder; Alzheimer's disease;

KW Parkinson's disease; cognitive disorder; schizophrenia; asthma;  
 KW skin disorder; psoriasis; sepsis; diabetes; diabetes; atherosclerosis;  
 KW cardiovascular disorder; angiogenic disorder; kidney disorder;  
 KW gastrointestinal disorder; pregnancy-related disorder; gene therapy;  
 KW endocrine disorder; infection; wound healing; vulnary; cell culture;  
 KW chemotaxis; food additive; binding partner identification.  
 XX Homo sapiens.  
 XX  
 PN WO200136432-A2.  
 XX  
 PD 25-MAY-2001.  
 XX  
 PF 15-NOV-2000; 2000WO-US011162.  
 XX  
 XX 19-NOV-1999; 99US-0166415P.  
 PR 30-JUN-2000; 2000US-0215138P.  
 XX  
 XX (HUMA-) HUMAN GENOME SCI INC.  
 PA  
 XX Ruben SM, Komatsoulis GA, Baker KP, Young PE;  
 PI WPI; 2001-343793/36.  
 XX  
 DR Isolated nucleic acid molecule encoding a human secreted protein is used  
 PT in preventing, treating or ameliorating a medical condition.  
 PT  
 XX Disclosure; Page 470-472; 509pp; English.  
 PS  
 XX  
 XX  
 CC AAD08488-AAD08529 represent cDNAs corresponding to 18 human secreted  
 CC protein genes, and AAE04199-AAE04239 represent the proteins they encode.  
 CC AAE04240-AAE04297 represent human secreted protein fragments or variants.  
 CC The secreted proteins and their genes are useful for preventing, treating  
 CC or ameliorating medical conditions, e.g., by protein or gene therapy.  
 CC Pathological conditions can be diagnosed by determining the amount of the  
 CC new protein in a sample or by determining the presence of mutations in  
 CC based on the tissues in which they are most highly expressed, and include  
 CC developing products for the diagnosis or treatment of proliferative  
 CC disorders, cancer, tumours, foetal and developmental abnormalities,  
 CC haematopoietic disorders, diseases of the immune system, AIDS, autoimmune  
 CC diseases (e.g., rheumatoid arthritis), inflammation, allergies,  
 CC neurological disorders (e.g., Alzheimer's disease, Parkinson's disease),  
 CC cognitive disorders, schizophrenia, asthma, skin disorders (e.g.,  
 CC psoriasis), sepsis, diabetes, atherosclerosis, gastrointestinal disorders,  
 CC angiogenic disorders, kidney disorders, endocrine disorders, and infections. The  
 CC pregnancy-related disorders, endocrine disorders, and infections. The  
 CC proteins can also be used to aid wound healing and epithelial cell  
 CC proliferation, to prevent skin ageing due to sunburn, to maintain organs  
 CC before transplantation, for supporting cell culture of primary tissues,  
 CC to regenerate tissues, to identify their cognate ligands or binding  
 CC partners, and in chemotaxis, and can be used as a food additive or  
 CC preservative to modify storage properties. Antibodies specific for a  
 CC protein of the invention can be used in alleviating symptoms associated  
 CC with the disorders mentioned above, and in diagnostic immunoassays e.g.,  
 CC radioimmunoassay or enzyme linked immunosorbent assay (ELISA). The  
 CC present sequence represents a human secreted protein fragment referred to  
 CC in the disclosure of the invention  
 XX  
 SQ

Query Match 53.0%; Score 390; DB 4; Length 592;  
 Best Local Similarity 99.7%; Pred. No. 0;  
 Matches 590; Conservative 0; Mismatches 2; Indels 0; Gaps 0;  
 QY 145 PNYLQVAYFKINKLEKAVAAAHFTFFVGNPEHMEMQNDLYQTMGSKGKADFQDLTQ 204  
 DB 1 PNYLQVAYFKINKLEKAVAAAHFTFFVGNPEHMEMQNDLYQTMGSKGKADFQDLTQ 60  
 QY 205 HMOEFLRLGLVRLYSEQPOEAVPHLEALQEFVAYEECRALCEGPDYDGYNYLEYNADL 264  
 DB 61 HMOEFLRLGLVRLYSEQPOEAVPHLEALQEFVAYEECRALCEGPDYDGYNYLEYNADL 120

QY 265 FOAITDHYIQLVCKQNCVTELASHPREKPFDFLPSHYNLYQFAYYNGYNTQTQGECA 324  
DB 121 FOAITDHYIQLVCKQNCVTELASHPREKPFDFLPSHYNLYQFAYYNGYNTQTQGECA 180  
QY 325 KTYLLFFPNDEVMQNLAYYAAVLGEEHTRSIGPRESAKYRQSRLLLEKELFFAYDVFG 384  
DB 181 KTYLLFFPNDEVMQNLAYYAAVLGEEHTRSIGPRESAKYRQSRLLLEKELFFAYDVFG 240  
QY 385 IPFVDPDSWTPPEVPIPKRLOEKOKSERETAVRISOEIGNLMKEIETLVEEKTESLDVSR 444  
DB 241 IPFVDPDSWTPPEVPIPKRLOEKOKSERETAVRISOEIGNLMKEIETLVEEKTESLDVSR 300  
QY 445 LTRGGGLLYEGISLTWNKLLNGYQVRVMDGVISDHECQELQRLTNVAATSGGVRGQT 504  
DB 301 LTRGGGLLYEGISLTWNKLLNGYQVRVMDGVISDHECQELQRLTNVAATSGGVRGQT 360  
QY 505 SPHTPNKFKYGVTVFKALKLQGEKVPLOSAHLYNNYVTEKVRIMESYFRLDTPLYESYS 564  
DB 361 SPHTPNKFKYGVTVFKALKLQGEKVPLOSAHLYNNYVTEKVRIMESYFRLDTPLYESYS 420  
QY 565 HLVCRTAIEEQAERKDDSHPHVNDNCILNAETLVCVKEPPAYTFRDYSAILYNGDFDG 624  
DB 421 HLVCRTAIEEQAERKDDSHPHVNDNCILNAETLVCVKEPPAYTFRDYSAILYNGDFDG 480  
QY 625 GNFTFTDLAKTVTAEVQPCGRVAVGSSGTENPHGVKAVTRGQRCALWFLDPRHSE 684  
DB 481 GNFTFTDLAKTVTAEVQPCGRVAVGSSGTENPHGVKAVTRGQRCALWFLDPRHSE 540  
QY 685 RDRVQADDLVQMLFSPREMDLSQEPDLAQCGPPEPAQESLSGSESKPKDEL 736  
DB 541 RDRVQADDLVQMLFSPREMDLSQEPDLAQCGPPEPAQESLSGSESKPKDEL 592  
RESULT 8  
ABG66709  
ID ABG66709 standard; protein; 806 AA.  
AC ABG66709;  
XX  
XX 30-AUG-2002 (first entry)  
XX  
XX Human novel polypeptide #44.  
XX  
XX Human; inflammatory condition; shock; sepsis; immune response; cancer;  
XX wound healing; central nervous system disease; haematopoiesis;  
XX peripheral nervous system disease; amyotrophic lateral sclerosis; tendon;  
XX myeloid cell disorder; lymphoid cell disorder; platelet disorder; bone;  
XX cartilage; ligament; nerve tissue; ulcer; osteoporosis; osteoarthritis;  
XX bone degenerative disorder; periodontal disease; reperfusion injury;  
XX lung fibrosis; liver fibrosis; autoimmune disorder; bacterial infection;  
XX allergic condition; thrombolysis; thrombosis; coagulation disorder;  
XX fungal infection.  
XX  
XX Homo sapiens.  
XX  
XX WO200244340-A2.  
XX  
XX 06-JUN-2002.  
XX  
XX 30-NOV-2001; 2001WO-US047004.  
XX  
XX 30-NOV-2000; 2000US-00728952.  
XX  
XX (HYSE-) HYSEQ INC.  
XX  
XX Tang YT, Goodrich RW, Liu C, Zhou P, Asundi V, Wang J, Wang D;  
XX Yanazaki V, Ujwal ML, Drmanac RT;  
XX  
XX WPI; 2002-508509/54.  
XX  
XX N-PSDB; ABK94933.  
XX  
XX Novel nucleic acids and polypeptides for diagnosis, treatment of  
XX inflammatory, autoimmune, nervous system, myeloid or lymphoid cell

PT disorders, cancer and promoting wound healing.  
XX Claim 10; Page 616-617; 672pp; English.  
PS  
XX The invention relates to human novel polynucleotides and associated  
XX polypeptides. The polynucleotides and polypeptides are useful for  
XX treating inflammatory conditions such as arthritis, nephritis, Crohn's  
XX disease, ischaemia-reperfusion injury, shock, sepsis, immune responses  
XX and cancer and for promoting wound healing. The sequences are used to  
XX induce the proliferation of neural cells and regeneration of nerve and  
XX brain tissue, and are useful for the treatment of central and peripheral  
XX nervous system diseases and neuropathies, such as Alzheimer's disease,  
XX Parkinson's disease, Huntington's disease and amyotrophic lateral  
XX sclerosis. The sequences are involved in chemotactic or chemokinetic  
XX activity, regulation of haematopoiesis, treatment of myeloid or lymphoid  
XX cell disorders and platelet disorders such as thrombocytopenia,  
XX regeneration of bone, cartilage, tendon, ligament and/or nerve tissue  
XX growth, tissue repair, healing of burns, incisions, ulcers, treatment of  
XX osteoporosis, osteoarthritis, bone degenerative disorders and periodontal  
XX disease. The sequences of the invention are also useful for gut  
XX protection or regeneration and treatment of lung or liver fibrosis,  
XX including severe combined immunodeficiency (SCID), bacterial or fungal  
XX infections, autoimmune disorders e.g. multiple sclerosis and myasthenia  
XX gravis, allergic conditions such as asthma, thrombolysis or thrombosis  
XX and coagulation disorders. Sequences ABG66666-ABG66758 represent human  
XX novel polypeptides of the invention  
SQ Sequence 806 AA;  
Query Match 51.6%; Score 380; DB 5; Length 806;  
Best Local Similarity 99.7%; Pred. No. 0;  
Matches 580; Conservative 0; Mismatches 2; Indels 0; Gaps 0;  
QY 155 KINKLEKAVAAHTFFVGNPEHMQOQLDYQTMGSKVEADFKOLETQPHQOEFLGVR 214  
DB 225 KINKLEKAVAAHTFFVGNPEHMQOQLDYQTMGSKVEADFKOLETQPHQOEFLGVR 284  
QY 215 LYSESQPOBAVPHLEAALQEQYFVAYEECRALCEGPDYDGVNLYEYNADLFOAITDHYIQ 274  
DB 285 LYSESQPOBAVPHLEAALQEQYFVAYEECRALCEGPDYDGVNLYEYNADLFOAITDHYIQ 344  
QY 275 VLNCQNCVTELASHPREKPFDFLPSHYNLYQFAYYNGYNTQTQGECAKTYLLFFPND 334  
DB 345 VLNCQNCVTELASHPREKPFDFLPSHYNLYQFAYYNGYNTQTQGECAKTYLLFFPND 404  
QY 335 EVMNQNLAYYAAVLGEEHTRSIGPRESAKYRQSRLLLEKELFFAYDVFGIPFVDPDSWT 394  
DB 405 EVMNQNLAYYAAVLGEEHTRSIGPRESAKYRQSRLLLEKELFFAYDVFGIPFVDPDSWT 464  
QY 395 PEEVTPKRLQEKOKSERETAVRISOEIGNLMKEIETLVEEKTESLDYSRLTREGGPLLY 454  
DB 465 PEEVTPKRLQEKOKSERETAVRISOEIGNLMKEIETLVEEKTESLDYSRLTREGGPLLY 524  
QY 455 EGISLTWNKLLNGYQVRVMDGVISDHECQELQRLTNVAATSGDGYRGQTSPTTNEKPY 514  
DB 525 EGISLTWNKLLNGYQVRVMDGVISDHECQELQRLTNVAATSGDGYRGQTSPTTNEKPY 584  
QY 515 GVTVPKALKLQGEKVPLOSAHLYNNYVTEKVRIMESYFRLDTPLYFSYSHLVCTAIEE 574  
DB 585 GVTVPKALKLQGEKVPLOSAHLYNNYVTEKVRIMESYFRLDTPLYFSYSHLVCTAIEE 644  
QY 575 VQAEKDDSHPHVNDNCILNAETLVCVKEPPAYTFRDYSAILYNGDFDGGNFYFTELDA 634  
DB 645 VQAEKDDSHPHVNDNCILNAETLVCVKEPPAYTFRDYSAILYNGDFDGGNFYFTELDA 704  
QY 635 KTVTAEVQPCGRVAVGSSGTENPHGVKAVTRGQRCALWFLDPRHSEDRVQADDLV 694  
DB 705 KTVTAEVQPCGRVAVGSSGTENPHGVKAVTRGQRCALWFLDPRHSEDRVQADDLV 764  
QY 695 KMLFSPREMDLSQEPDLAQCGPPEPAQESLSGSESKPKDEL 736  
DB 765 KMLFSPREMDLSQEPDLAQCGPPEPAQESLSGSESKPKDEL 806

RESULT 9  
 ABG66686  
 ID ABG66686 standard; protein; 806 AA.  
 XX  
 AC ABG66686;  
 XX  
 DT 30-AUG-2002 (first entry)  
 XX  
 DE Human novel polypeptide #21.  
 XX  
 KW Human; inflammatory condition; shock; sepsis; immune response; cancer;  
 KW wound healing; central nervous system disease; haematopoiesis;  
 KW peripheral nervous system disease; amyotrophic lateral sclerosis; tendon;  
 KW myeloid cell disorder; lymphoid cell disorder; platelet disorder; bone;  
 KW cartilage; ligament; nerve tissue; ulcer; osteoporosis; osteoarthritis;  
 KW bone degenerative disorder; periodontal disease; reperfusion injury;  
 KW lung fibrosis; liver fibrosis; autoimmune disorder; bacterial infection;  
 KW allergic condition; thrombolytic; thrombosis; coagulation disorder;  
 KW fungal infection.  
 XX  
 OS Homo sapiens.  
 XX  
 PN WO200244340-A2.  
 XX  
 PD 06-JUN-2002.  
 XX  
 PF 30-NOV-2001; 2001WO-US047004.  
 XX  
 PR 30-NOV-2000; 2000US-00728952.  
 XX  
 PA (HYSE-) HYSEQ INC.  
 XX  
 PI Tang YT, Goodrich RW, Liu C, Zhou P, Asundi V, Wang J, Wang D;  
 PI Yamazaki V, Ujwal ML, Drmanac RT;  
 XX  
 DR N-PSDB; ABK94910.  
 XX  
 PT Novel nucleic acids and polypeptides for diagnosis, treatment of  
 PT inflammatory, autoimmune, nervous system, myeloid or lymphoid cell  
 PT disorders, cancer and promoting wound healing.  
 XX  
 PS Claim 10; Page 584-585; 672pp; English.  
 XX  
 CC The invention relates to human novel polynucleotides and associated  
 CC polypeptides. The polynucleotides and polypeptides are useful for  
 CC treating inflammatory conditions such as arthritis, nephritis, Crohn's  
 CC disease, ischaemia-reperfusion injury, shock, sepsis, immune responses  
 CC and cancer and for promoting wound healing. The sequences are used to  
 CC induce the proliferation of neural cells and regeneration of nerve and  
 CC brain tissue, and are useful for the treatment of central and peripheral  
 CC nervous system diseases and neuropathies, such as Alzheimer's disease,  
 CC Parkinson's disease, Huntington's disease and amyotrophic lateral  
 CC sclerosis. The sequences are involved in chemotactic or chemokinetic  
 CC activity, regulation of haematopoiesis, treatment of myeloid or lymphoid  
 CC cell disorders and platelet disorders such as thrombocytopenia,  
 CC regeneration of bone, cartilage, tendon, ligament and/or nerve tissue  
 CC growth, tissue repair, healing of burns, incisions, ulcers, treatment of  
 CC osteoporosis, osteoarthritis, bone degenerative disorders and periodontal  
 CC disease. The sequences of the invention are also useful for gut  
 CC protection or regeneration and treatment of lung or liver fibrosis,  
 CC reperfusion injury in various tissues, immune deficiencies and disorders  
 CC including severe combined immunodeficiency (SCID), bacterial or fungal  
 CC infections, autoimmune disorders e.g. multiple sclerosis and myasthenia  
 CC gravis, allergic conditions such as asthma, thrombolytic or thrombotic  
 CC and coagulation disorders. Sequences ABG66666-ABG66758 represent human  
 CC novel polypeptides of the invention.  
 XX  
 SQ Sequence 806 AA;

Best Local Similarity 99.7%; Pred. No. 0;		Matches 580; Conservative 0; Mismatches 2; Indels 0; Gaps 0;	
QY	155	KINKLEKAVAAAHFTFVGNPEHMEQNLDDYQTMGSKVEADFKDLETPQPMQERLQVR	214
DB	225	KINKLEKAVAAAHFTFVGNPEHMEQNLDDYQTMGSKVEADFKDLETPQPMQERLQVR	284
QY	215	LYSEBQPOEAVPHLEAALQOEYFVAYEECRALCEGVDYDGNLYNLEYNADLQAITDHIQ	274
DB	285	LYSEBQPOEAVPHLEAALQOEYFVAYEECRALCEGVDYDGNLYNLEYNADLQAITDHIQ	344
QY	275	VLNCKQNCVTELASHPSREKPFEDFLPSHNYNQFAYNIGNYTOAGCAKTYLLFFPND	334
DB	345	VLNCKQNCVTELASHPSREKPFEDFLPSHNYNQFAYNIGNYTOAGCAKTYLLFFPND	404
QY	335	EYVQNQLAYAAAMIGEEHTRSGPRESAKYRQSRLLKELLFFAYDVGIPFVDPDSMT	394
DB	405	EYVQNQLAYAAAMIGEEHTRSGPRESAKYRQSRLLKELLFFAYDVGIPFVDPDSMT	464
QY	395	PEEVIPKRLQEKQKSERETAVRAISOEIGNLMKEIETLVEEKTESLDVSRLTREGGPLY	454
DB	465	PEEVIPKRLQEKQKSERETAVRAISOEIGNLMKEIETLVEEKTESLDVSRLTREGGPLY	524
QY	455	EGISLTMSKLLNGYQRYVMGVIISDHCEQELQRLTNVAATSGDGYRGQTSPTHTNEKPY	514
DB	525	EGISLTMSKLLNGYQRYVMGVIISDHCEQELQRLTNVAATSGDGYRGQTSPTHTNEKPY	584
QY	515	GTVTFKALKIGQEGKVPQSAHLYNNVTEKVRIMESYFRDLTPLYFSYSHLVCRTAEE	574
DB	585	GTVTFKALKIGQEGKVPQSAHLYNNVTEKVRIMESYFRDLTPLYFSYSHLVCRTAEE	644
QY	575	VQAEKDDSHPVHVDNCLINAEFLVCVKEPPAYTFRDYSAILYLNGDFDGGNFYFTELDA	634
DB	645	VQAEKDDSHPVHVDNCLINAEFLVCVKEPPAYTFRDYSAILYLNGDFDGGNFYFTELDA	704
QY	635	KTVTAEVQPCGRVGFSSGTENPHGVKAVTRGQRCALAMFTLDPHRSERDRVQADDLV	694
DB	705	KTVTAEVQPCGRVGFSSGTENPHGVKAVTRGQRCALAMFTLDPHRSERDRVQADDLV	764
QY	695	KMLFSPDEMDLSQEQPLDAQQGPPEPAQESLSGSESKPKDEL	736
DB	765	KMLFSPDEMDLSQEQPLDAQQGPPEPAQESLSGSESKPKDEL	806

RESULT 10  
 AAB36391  
 ID AAB36391 standard; protein; 363 AA.  
 XX  
 AC AAB36391;  
 XX  
 DT 27-FEB-2001 (first entry)  
 XX  
 DE Human tumour suppressor Gros1-L protein SEQ ID NO:2.  
 XX  
 KW Tumour suppressor; Gros1-L; Gros1-S; cell proliferation; regulation;  
 KW cancer; cytostatic; gene therapy.  
 XX  
 OS Homo sapiens.  
 XX  
 PN WO200065047-A1.  
 XX  
 PD 02-NOV-2000.  
 XX  
 PF 26-APR-2000; 2000WO-JP002731.  
 XX  
 PR 26-APR-1999; 99JP-00118806.  
 XX  
 PA (CHUG-) CHUGAI RES INST MOLECULAR MEDICINE INC.  
 XX  
 PI Wadhwa R, Sugihara T, Yoshida A;  
 XX  
 DR WPI; 2000-687340/67.



XX Full-length tumor suppressor genes encoding Grosi-L, Grosi-S and mouse  
PT homologs participating in regulation of cell proliferation, useful in  
PT development of preventives and remedies of cancer.  
XX  
XX Claim 1; Page 67-69; 114pp; Japanese.  
XX  
XX The present sequence represents the human tumour suppressor designated  
CC Grosi-L. Grosi-L and Grosi-S have cytostatic activity and can be used in  
CC gene therapy. Grosi-L and Grosi-S genes are useful in the development of  
CC drugs used to treat and prevent cancer  
XX  
SQ Sequence 363 AA;  
  
Query Match 48.9%; Score 360; DB 3; Length 363;  
Best Local Similarity 100.0%; Pred. No. 0;  
Matches 360; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
  
Qy 1 MAVZALKLTTLLAVVAAASQAQVESEAGWGVTPLLFAEGTAAYARGDWPVLSMER 60  
Db 1 MAVZALKLTTLLAVVAAASQAQVESEAGWGVTPLLFAEGTAAYARGDWPVLSMER 60  
  
Qy 61 ALRSRAALRALRLRCRTQCAADFPWELDPWSPSPAQASGAGALRDLSPFGGLLRAACL 120  
Db 61 ALRSRAALRALRLRCRTQCAADFPWELDPWSPSPAQASGAGALRDLSPFGGLLRAACL 120  
  
Qy 121 RRCLGPPAAHSLSEEMELEFRKRSPPNYLQVAYFKINKLEKAVAAAHFTFFVGNPEHMEMQ 180  
Db 121 RRCLGPPAAHSLSEEMELEFRKRSPPNYLQVAYFKINKLEKAVAAAHFTFFVGNPEHMEMQ 180  
  
Qy 181 QNLDDYQTMGSGVKEADFKDLETQPHMOEFLGRLVLYSEEQPOEAVPHLEALQEFYFVAYE 240  
Db 181 QNLDDYQTMGSGVKEADFKDLETQPHMOEFLGRLVLYSEEQPOEAVPHLEALQEFYFVAYE 240  
  
Qy 241 ECRALCEGPDYDGYNLYEYNADLFOAITDHYIQLVNLCKQNCVTELASHPSREKPPDFL 300  
Db 241 ECRALCEGPDYDGYNLYEYNADLFOAITDHYIQLVNLCKQNCVTELASHPSREKPPDFL 300  
  
Qy 301 PSHYNYLQFAYNIGNYTOAGECAKTYLLFFPNDEVNQNLAAYAAAMLGEHSTRSIGPRE 360  
Db 301 PSHYNYLQFAYNIGNYTOAGECAKTYLLFFPNDEVNQNLAAYAAAMLGEHSTRSIGPRE 360  
  
RESULT 11  
ID AAB94482  
XX AAB94482 standard; protein; 804 AA.  
XX  
XX AAB94482;  
XX  
XX 26-JUN-2001 (first entry)  
XX  
XX Human protein sequence SEQ ID NO:15160.  
XX  
XX Human; primer; detection; diagnosis; antisense therapy; gene therapy.  
XX  
XX Homo sapiens.  
XX  
XX EP1074617-A2.  
XX  
XX 07-FEB-2001.  
XX  
XX 28-JUL-2000; 2000EP-00116126.  
XX  
XX 29-JUL-1999; 99JP-00248036.  
XX  
XX 27-AUG-1999; 99JP-00300253.  
XX  
XX 11-JAN-2000; 2000JP-00118776.  
XX  
XX 02-MAY-2000; 2000JP-00183767.  
XX  
XX 09-JUN-2000; 2000JP-00241899.  
XX  
XX (HELI-) HELIX RES INST.  
XX  
XX Ota T, Isogai T, Nishikawa T, Hayashi K, Saito K, Yamamoto J;  
PI Ishii S, Sugiyama T, Wakamatsu A, Nagai K, Otsuki T;

XX WPI; 2001-318749/34.  
XX  
XX Primer sets for synthesizing polynucleotides, particularly the 5602 full-  
PT length cDNAs defined in the specification, and for the detection and/or  
PT diagnosis of the abnormality of the proteins encoded by the full-length  
PT cDNAs.  
XX  
XX Claim 8; SEQ ID NO 15160; 2537pp + Sequence Listing; English.  
XX  
XX The present invention describes primer sets for synthesizing 5602 full-  
CC length cDNAs defined in the specification. Where a primer set comprises:  
CC (a) an oligo-dT primer and an oligonucleotide complementary to the  
CC complementary strand of a polynucleotide which comprises one of the 5602  
CC nucleotide sequences defined in the specification, where the  
CC oligonucleotide comprises at least 15 nucleotides; or (b) a combination  
CC of an oligonucleotide comprising a sequence complementary to the  
CC complementary strand of a polynucleotide which comprises a 5'-end  
CC sequence and an oligonucleotide comprising a sequence complementary to a  
CC polynucleotide which comprises a 3'-end sequence, where the  
CC oligonucleotide comprises at least 15 nucleotides and the combination of  
CC the 5'-end sequence/3'-end sequence is selected from those defined in the  
CC specification. The primer sets can be used in antisense therapy and in  
CC gene therapy. The primers are useful for synthesizing polynucleotides,  
CC particularly full-length cDNAs. The primers are also useful for the  
CC detection and/or diagnosis of the abnormality of the proteins encoded by  
CC the full-length cDNAs. The primers allow obtaining of the full-length  
CC cDNAs easily without any specialised methods. AAH03166 to AAH13628 and  
CC AAH13633 to AAH18742 represent human cDNA sequences; AAB92446 to AAB95893  
CC represent human amino acid sequences; and AAH13629 to AAH13632 represent  
CC oligonucleotides, all of which are used in the exemplification of the  
CC present invention  
XX  
SQ Sequence 804 AA;  
  
Query Match 38.2%; Score 281; DB 4; Length 804;  
Best Local Similarity 99.4%; Pred. No. 47e-262;  
Matches 681; Conservative 0; Mismatches 4; Indels 0; Gaps 0;  
  
Qy 1 MAVRALKLTLLAVVAAASQAQVESEAGWGVTPLLFAEGTAAYARGDWPVLSMER 60  
Db 1 MAVRALKLTLLAVVAAASQAQVESEAGWGVTPLLFAEGTAAYARGDWPVLSMER 60  
  
Qy 61 ALRSRAALRALRLRCRTQCAADFPWELDPWSPSPAQASGAGALRDLSPFGGLLRAACL 120  
Db 61 ALRSRAALRALRLRCRTQCAADFPWELDPWSPSPAQASGAGALRDLSPFGGLLRAACL 120  
  
Qy 121 RRCLGPPAAHSLSEEMELEFRKRSPPNYLQVAYFKINKLEKAVAAAHFTFFVGNPEHMEMQ 180  
Db 121 RRCLGPPAAHSLSEEMELEFRKRSPPNYLQVAYFKINKLEKAVAAAHFTFFVGNPEHMEMQ 180  
  
Qy 181 QNLDDYQTMGSGVKEADFKDLETQPHMOEFLGRLVLYSEEQPOEAVPHLEALQEFYFVAYE 240  
Db 181 QNLDDYQTMGSGVKEADFKDLETQPHMOEFLGRLVLYSEEQPOEAVPHLEALQEFYFVAYE 240  
  
Qy 241 ECRALCEGPDYDGYNLYEYNADLFOAITDHYIQLVNLCKQNCVTELASHPSREKPPDFL 300  
Db 241 ECRALCEGPDYDGYNLYEYNADLFOAITDHYIQLVNLCKQNCVTELASHPSREKPPDFL 300  
  
Qy 301 PSHYNYLQFAYNIGNYTOAGECAKTYLLFFPNDEVNQNLAAYAAAMLGEHSTRSIGPRE 360  
Db 301 PSHYNYLQFAYNIGNYTOAGECAKTYLLFFPNDEVNQNLAAYAAAMLGEHSTRSIGPRE 360  
  
Qy 361 SAKERYORSLLLEKELLFFAYDVFGLPFVDPDPSWTPEEVIKPKLOEKQKSERETAVRISOE 420  
Db 361 SAKERYORSLLLEKELLFFAYDVFGLPFVDPDPSWTPEEVIKPKLOEKQKSERETAVRISOE 420  
  
Qy 421 IGNLMKEIETLVEEKTKESLDVSRLTREGGPLLLEGISLTMTNSKLANGYQVRVMDGVISD 480  
Db 421 IGNLMKEIETLVEEKTKESLDVSRLTREGGPLLLEGISLTMTNSKLANGYQVRVMDGVISD 480  
  
Qy 481 HECQELQRLTNVAATSGDGYRGQTSPHTNEKFGYVTVFKALKLGGKVPCLSAHLNYN 540  
Db 481 HECQELQRLTNVAATSGDGYRGQTSPHTNEKFGYVTVFKALKLGGKVPCLSAHLNYN 540

Db 481 HECQELQRLTNVAATSGDGYRGQTSPTNEKFGYGVTVFKALKLGQEGKVPLOSAHLYN 540

QY 541 VTEKVRIMESYFRDLTPLYFSYSHLVCRTAIEEVOAERKDDSHPVHVDNCILNAETLVC 600

Db 541 VTEKVRIMESYFRDLTPLYFSYSHLVCRTAIEEVOAERKDDSHPVHVDNCILNAETLVC 600

QY 601 VKEPPAYTRDYSAIYLNGDFDGGNFYFTELDKTVTAEVQPOCGRAVGFSSGTENPHG 660

Db 601 VKEPPAYTRDYSAIYLNGDFDGGNFYFTELDKTVTAEVQPOCGRAVGFSSGTENPHG 660

QY 661 VKAVTRGQRCALWFTLDPHRSER 685

Db 661 VKAVTRGQRCALWFTLDPHRSER 685

RESULT 12

AAU09860

ID AAU09860 standard; protein; 708 AA.

XX

AC AAU09860;

XX

DT 26-FEB-2002 (first entry)

XX

DE Novel human secreted protein #1.

XX

KW Secreted protein; cytostatic; immunosuppressive; vulnery; vaccine;

KW antiinflammatory; neuroprotective; nephrotropic; cardiovascular; human;

KW cancer; autoimmune disease; wound healing disorder; infectious;

KW haematopoietic disorder; inflammatory disorder; infertility;

KW neurological disease; psychiatric disease; cardiovascular disease;

KW respiratory disease; renal; gastrointestinal.

XX

OS Homo sapiens.

XX

PN WO200179454-A1.

XX

PD 25-OCT-2001.

XX

PF 11-APR-2001; 2001WO-US011797.

XX

PR 13-APR-2000; 2000US-0196603P.

PR 24-APR-2000; 2000US-0199417P.

XX

PA (SMIK ) SMITHKLINE BEECHAM CORP.

PA (SMIK ) SMITHKLINE BEECHAM PLC.

XX

PI Agarwal P, Murdoch PR, Rizvi SK, Smith RF, Xiang Z;

DR WPI; 2002-061975/08.

DR N-PSDB; AAS17572.

XX

PT New secreted proteins or polypeptides, useful for treating e.g. cancer,

PT autoimmune diseases, wound healing disorder, infections, haematopoietic

PT disorders, inflammatory disorders, infertility, cancer.

XX

PS Claim 1; Page 52-54; 92pp; English.

XX

CC The invention relates to an isolated novel secreted polypeptide (I) and

CC polynucleotide (II). (I) and (II) are useful for treating cancer,

CC autoimmune diseases, wound healing disorder, infections, haematopoietic

CC disorders, inflammatory disorders, infertility, neurological and

CC psychiatric diseases, cardiovascular diseases, respiratory diseases,

CC renal diseases, or gastrointestinal diseases. These may also be used to

CC treat diseases, abnormalities and disorders caused by abnormal

CC expression, production, function and/or metabolism of the genes, as

CC vaccines for inducing immunological response in a mammal, and in

CC screening methods for detecting the effect of added compounds on the

CC production of mRNA and polypeptide in cells. The polypeptides can be used

CC as immunogens to produce antibodies immunospecific for the polypeptides,

CC and to identify membrane-bound or soluble receptors. The polynucleotides

CC may be used as diagnostic reagents, in chromosome localisation studies,

CC and in tissue expression studies. The present sequence represents the

CC amino acid sequence of novel human secreted protein #1

XX SQ Sequence 708 AA;

Query Match 37.4%; Score 275; DB 5; Length 708;

Best Local Similarity 99.7%; Pred. No. 2.6e-256;

Matches 375; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 361 SAKYRORSLLKEILLFFAYDVREGIPFVDPDSDWTPEEIPKLOEKOKSERETAVRISQE 420

Db 333 SAKYRORSLLKEILLFFAYDVREGIPFVDPDSDWTPEEIPKLOEKOKSERETAVRISQE 392

QY 421 IGNLMKEIETLVEEKTKEISLDVSRLTREGGPLLVEGIGISLTNNKLLNGYQVRVMDGVISD 480

Db 393 IGNLMKEIETLVEEKTKEISLDVSRLTREGGPLLVEGIGISLTNNKLLNGYQVRVMDGVISD 452

QY 481 HECQELQRLTNVAATSGDGYRGQTSPTNEKFGYGVTVFKALKLGQEGKVPLOSAHLYN 540

Db 453 HECQELQRLTNVAATSGDGYRGQTSPTNEKFGYGVTVFKALKLGQEGKVPLOSAHLYN 512

QY 541 VTEKVRIMESYFRDLTPLYFSYSHLVCRTAIEEVOAERKDDSHPVHVDNCILNAETLVC 600

Db 513 VTEKVRIMESYFRDLTPLYFSYSHLVCRTAIEEVOAERKDDSHPVHVDNCILNAETLVC 572

QY 601 VKEPPAYTRDYSAIYLNGDFDGGNFYFTELDKTVTAEVQPOCGRAVGFSSGTENPHG 660

Db 573 VKEPPAYTRDYSAIYLNGDFDGGNFYFTELDKTVTAEVQPOCGRAVGFSSGTENPHG 632

QY 661 VKAVTRGQRCALWFTLDPHRSERDRVQADDLVKMLFSPEEMDLSQSPQLDAQOGPPPEP 720

Db 633 VKAVTRGQRCALWFTLDPHRSERDRVQADDLVKMLFSPEEMDLSQSPQLDAQOGPPPEP 692

QY 721 AQESLSGSESKEPKDEL 736

Db 693 AQESLSGSESKEPKDEL 708

RESULT 13

AAE04246

ID AAE04246 standard; protein; 337 AA.

XX

AC AAE04246;

XX

DT 09-AUG-2001 (first entry)

XX

DE Human gene 2 encoded secreted protein fragment, SEQ ID NO:103.

XX

KW Human; secreted protein; proliferative disorder; cancer; tumour;

KW foetal abnormality; developmental abnormality; haematopoietic disorder;

KW immune system disorder; AIDS; autoimmune disease; rheumatoid arthritis;

KW inflammation; allergy; neurological disorder; Alzheimer's disease;

KW Parkinson's disease; cognitive disorder; schizophrenia; asthma;

KW skin disorder; psoriasis; sepsis; diabetes; kidney disorder;

KW cardiovascular disorder; angiogenic disorder; kidney disorder;

KW gastrointestinal disorder; pregnancy-related disorder; gene therapy;

KW endocrine disorder; infection; wound healing; vulnery; cell culture;

KW chemotaxis; food additive; binding partner identification.

XX

OS Homo sapiens.

XX

PN WO200136432-A2.

XX

PD 25-MAY-2001.

XX

PF 15-NOV-2000; 2000WO-US031162.

XX

PR 19-NOV-1999; 99US-0166415P.

PR 30-JUN-2000; 2000US-0215136P.

XX

PA (HUMA-) HUMAN GENOME SCI INC.

XX

PI Ruben SM, Komatsoulis GA, Baker KP, Young PE;

DR WPI; 2001-343793/36.



XX Isolated nucleic acid molecule encoding a human secreted protein is used  
PT in preventing, treating or ameliorating a medical condition.  
XX  
XX Disclosure; Page 12; 509pp; English.  
XX  
XX AAD08488-AAD08529 represent cDNAs corresponding to 18 human secreted  
CC protein genes, and AAE04199-AAE04239 represent the proteins they encode.  
CC AAD04240-AAE04297 represent human secreted protein fragments or variants.  
CC The secreted proteins and their genes are useful for preventing, treating  
CC or ameliorating medical conditions, e.g., by protein or gene therapy.  
CC Pathological conditions can be diagnosed by determining the amount of the  
CC new protein in a sample or by determining the presence of mutations in  
CC the new genes. Specific uses are described for each of the 18 genes,  
CC based on the tissues in which they are most highly expressed, and include  
CC developing products for the diagnosis or treatment of proliferative  
CC disorders, cancer, tumors, foetal and developmental abnormalities,  
CC haematopoietic disorders, diseases of the immune system, AIDS, autoimmune  
CC diseases (e.g., rheumatoid arthritis), inflammation, allergies,  
CC neurological disorders (e.g., Alzheimer's disease, Parkinson's disease),  
CC cognitive disorders, schizophrenia, asthma, skin disorders (e.g.,  
CC psoriasis), sepsis, diabetes, atherosclerosis, cardiovascular disorders,  
CC angioecnic disorders, kidney disorders, gastrointestinal disorders,  
CC pregnancy-related disorders, endocrine disorders, and infections. The  
CC proteins can also be used to aid wound healing and epithelial cell  
CC proliferation, to prevent skin ageing due to sunburn, to maintain organs  
CC before transplantation, for supporting cell culture of primary tissues,  
CC to regenerate tissues, to identify their cognate ligands or binding  
CC partners, and in chemo-axis, and can be used as a food additive or  
CC preservative to modify storage properties. Antibodies specific for a  
CC protein of the invention can be used in alleviating symptoms associated  
CC with the disorders mentioned above, and in diagnostic immunoassays e.g.,  
CC radioimmunoassay or enzyme linked immunosorbent assay (ELISA). The  
CC present sequence represents a human secreted protein fragment referred to  
CC in the disclosure of the invention  
XX  
XX SQ Sequence 337 AA;

Query Match 36.3%; Score 267; DB 4; Length 337;  
Best Local Similarity 100.0%; Pred. No. 7.3e-249;  
Matches 267; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 470 QRVMDGVISDHECOELQRLTNVAATSGDGYRGQTSPTNEKFGYGVTVFKALKLGOEGK 529  
DB 71 QRVMDGVISDHECOELQRLTNVAATSGDGYRGQTSPTNEKFGYGVTVFKALKLGOEGK 130  
QY 530 VPLQSAHLYNNVTEKVRIMESYRLDTPLYFSYSHLVCRTAIEVQAEKDDSHPVHVD 589  
DB 131 VPLQSAHLYNNVTEKVRIMESYRLDTPLYFSYSHLVCRTAIEVQAEKDDSHPVHVD 190  
QY 590 NCILNAETLVCKVBPAYTFRDYSALLYLNGDFDGGNFYFTLDAKTVTAEVQPCGRAV 649  
DB 191 NCILNAETLVCKVBPAYTFRDYSALLYLNGDFDGGNFYFTLDAKTVTAEVQPCGRAV 250  
QY 650 GFSSGTENPHGVKAVTRGQRCALWFTLDRHSDRVDQADDLVKMLFSPPEMDLSQEQ 709  
DB 251 GFSSGTENPHGVKAVTRGQRCALWFTLDRHSDRVDQADDLVKMLFSPPEMDLSQEQ 310  
QY 710 PLDAQOQGPPEPAQESLSGSESKPKDEL 736  
DB 311 PLDAQOQGPPEPAQESLSGSESKPKDEL 337

RESULT 14  
ABP69128  
ID ABP69128 standard; protein; 706 AA.  
XX  
XX AC ABP69128;  
XX  
DT 20-JAN-2003 (first entry)  
XX  
DE Human polypeptide SEQ ID NO 1175.  
XX

KW Human; genome mapping; gene therapy; food supplement; virus; fungus;  
KW cell-proliferative disorder; neurodegenerative disease; bacterial;  
KW Parkinson's disease; Alzheimer's disease; autoimmune disease;  
KW multiple sclerosis; diabetes; genetic disorder; wound; burn; infection;  
KW arthritis; cystostatic; immunomodulator; neurotropic; dermatological;  
KW antiparkinsonian; antidiabetic; immunosuppressive; haematological;  
KW haemostatic; vulnary; fungicide; antibacterial; virucide; protozoacide;  
KW antiarthritic.  
XX Homo sapiens.  
OS  
XX  
XX WO200270539-A2.  
PN  
XX 12-SEP-2002.  
PD  
XX  
XX 05-MAR-2002; 2002WO-US005095.  
PF  
XX  
XX 05-MAR-2001; 2001US-00799451.  
PR  
XX (HYSE-) HYSEQ INC.  
PA  
XX Tang YT, Zhou P, Goodrich RW, Asundi V, Zhang J, Zhao QA, Ren F;  
PI Xue AJ, Yang Y, Ma Y, Yamazaki V, Chen R, Wang Z, Ghosh M;  
PI Wehrman T, Wang J, Wang D, Drmanac RT;  
XX WPI; 2002-759812/82.  
DR N-PSDB; ABZ11345.  
XX  
XX New polynucleotides comprising sequences assembled from expressed  
PT sequence tags (ESTs), useful for treating cell-proliferative, or platelet  
PT neurodegenerative, autoimmune, genetic, myeloid or lymphoid, or platelet  
PT or coagulation disorders.  
XX  
XX Claim 9; SEQ ID NO 1175; 1012pp + Sequence Listing; English.  
XX  
XX The invention relates to an isolated polynucleotide (I) comprising a  
CC nucleotide sequence selected from any of 948 sequences (ABZ11119-  
CC ABZ12066) or their mature protein coding portion, active domain coding  
CC protein or complementary sequences. The polynucleotides are useful for  
CC identifying expressed genes or for physical mapping of human genome. The  
CC encoded polypeptides (ABP68902-ABP69849) are useful as molecular weight  
CC markers, as a food supplement, for generating antibodies, in medical  
CC imaging, screening and diagnostic assays and for treating cell-  
CC proliferative disorders (cancer), neurodegenerative diseases (Parkinson's  
CC or Alzheimer's disease), autoimmune diseases (multiple sclerosis,  
CC diabetes, lupus) genetic disorders, myeloid or lymphoid disorders,  
CC platelet or coagulation disorders, wound, burns, incision, ulcers, liver  
CC or lung fibrosis, infections (bacterial, viral, fungal, parasitic),  
CC arthritis, etc. Note: The sequence data for this patent did not form part  
CC of the printed specification, but was obtained in electronic format  
CC directly from WIPO at ftp.wipo.int/pub/published\_pct\_sequences  
XX  
XX SQ Sequence 706 AA;

Query Match 33.2%; Score 244; DB 5; Length 706;  
Best Local Similarity 100.0%; Pred. No. 2.3e-226;  
Matches 244; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 493 AATSGDGYRGQTSPTNEKFGYGVTVFKALKLGOEGKVPLOSAHLYNNVTEKVRIMESY 552  
DB 463 AATSGDGYRGQTSPTNEKFGYGVTVFKALKLGOEGKVPLOSAHLYNNVTEKVRIMESY 522  
QY 553 FRLDTPLYFSYSHLVCRTAIEVQAEKDDSHPVHVDNCILNAETLVCKVPEPPAYTFRDY 612  
DB 523 FRLDTPLYFSYSHLVCRTAIEVQAEKDDSHPVHVDNCILNAETLVCKVPEPPAYTFRDY 582  
QY 613 SALLYLNGDFDGGNFYFTLDAKTVTAEVQPCGRAVGFSSGTENPHGVKAVTRGQRCAL 672  
DB 583 SALLYLNGDFDGGNFYFTLDAKTVTAEVQPCGRAVGFSSGTENPHGVKAVTRGQRCAL 642  
QY 673 ALWFTLDRHSDRVDQADDLVKMLFSPPEMDLSQEQPLDAQOQGPPEPAQESLSGSESKP 732  
DB 643 ALWFTLDRHSDRVDQADDLVKMLFSPPEMDLSQEQPLDAQOQGPPEPAQESLSGSESKP 702

QY 733 KDEL 736  
 DB 703 KDEL 706

RESULT 15  
 AA04200  
 ID AA04200 standard; protein; 359 AA.  
 AC AA04200;  
 XX  
 XX 09-AUG-2001 (first entry)  
 XX  
 DE Human gene 2 encoded secreted protein HUVFY29, SEQ ID NO:54.  
 XX  
 KW Human; secreted protein; proliferative disorder; cancer; tumour;  
 KW foetal abnormality; developmental abnormality; haematopoietic disorder;  
 KW immune system disorder; AIDS; autoimmune disease; rheumatoid arthritis;  
 KW inflammation; allergy; neurological disorder; Alzheimer's disease;  
 KW Parkinson's disease; cognitive disorder; schizophrenia; asthma;  
 KW skin disorder; psoriasis; sepsis; diabetes; atherosclerosis;  
 KW cardiovascular disorder; angiogenic disorder; kidney disorder;  
 KW gastrointestinal disorder; pregnancy-related disorder; gene therapy;  
 KW endocrine disorder; infection; wound healing; vulnerability; cell culture;  
 KW chemotaxis; food additive; binding partner identification.  
 XX  
 OS Homo sapiens.  
 XX  
 XX Key Location/Qualifiers  
 FT Peptide 1..18  
 FT Protein 19..359  
 FT /label= signal\_peptide  
 FT /note= "Mature secreted protein"  
 XX  
 PN WO200136432-A2.  
 XX  
 PD 25-MAY-2001.  
 XX  
 XX 15-NOV-2000; 2000WO-US031162.  
 XX  
 XX 19-NOV-1999; 99US-0166415P.  
 PR 30-JUN-2000; 2000US-0215136P.  
 XX  
 PA (HUMA-) HUMAN GENOME SCI INC.  
 XX  
 PI Ruben SM, Komatsoulis GA, Baker KP, Young PE;  
 XX  
 DR WPI; 2001-343793/36.  
 DR N-PSDB; AAD08489.  
 XX  
 PT Isolated nucleic acid molecule encoding a human secreted protein is used  
 PT in preventing, treating or ameliorating a medical condition.  
 XX  
 PS Claim 11; Page 431-433; 509pp; English.  
 XX  
 CC AAD08488-AAD08529 represent cDNAs corresponding to 18 human secreted  
 CC protein genes, and AA04199-AA04239 represent the proteins they encode.  
 CC AA04240-AA04297 represent human secreted protein fragments or variants.  
 CC The secreted proteins and their genes are useful for preventing, treating  
 CC or ameliorating medical conditions, e.g., by protein or gene therapy.  
 CC Pathological conditions can be diagnosed by determining the amount of the  
 CC new protein in a sample or by determining the presence of mutations in  
 CC the new genes. Specific uses are described for each of the 18 genes,  
 CC based on the tissues in which they are most highly expressed, and include  
 CC developing products for the diagnosis or treatment of proliferative  
 CC disorders, cancer, tumours, foetal and developmental abnormalities,  
 CC haematopoietic disorders, diseases of the immune system, AIDS, autoimmune  
 CC diseases (e.g., rheumatoid arthritis), inflammation, allergies,  
 CC neurological disorders (e.g., Alzheimer's disease, Parkinson's disease),  
 CC cognitive disorders, schizophrenia, asthma, skin disorders (e.g.,  
 CC psoriasis), sepsis, diabetes, atherosclerosis, cardiovascular disorders,  
 CC angiogenic disorders, kidney disorders, gastrointestinal disorders,

CC pregnancy-related disorders, endocrine disorders, and infections. The  
 CC proteins can also be used to aid wound healing and epithelial cell  
 CC proliferation, to prevent skin ageing due to sunburn, to maintain organs  
 CC before transplantation, for supporting cell culture of primary tissues,  
 CC to regenerate tissues, to identify their cognate ligands or binding  
 CC partners, and in chemotaxis, and can be used as a food additive or  
 CC preservative to modify storage properties. Antibodies specific for a  
 CC protein of the invention can be used in alleviating symptoms associated  
 CC with the disorders mentioned above, and in diagnostic immunoassays e.g.,  
 CC radioimmunoassay or enzyme linked immunosorbent assay (ELISA). The  
 CC present sequence represents a human secreted protein of the invention  
 XX  
 SQ Sequence 359 AA;  
 Query Match 29.8%; Score 219; DB 4; Length 359;  
 Best Local Similarity 99.7%; Pred. No. 1.8e-202;  
 Matches 319; Conservative 0; Mismatches 1; Indels 0; Gaps 0;  
 QY 1 MAVRALKLTLLAVVAASQAEVESEAGWMTDLLFAEGTAAYARGDWGVLSMER 60  
 DB 1 MAVRALKLTLLAVVAASQAEVESEAGWMTDLLFAEGTAAYARGDWGVLSMER 60  
 QY 61 ALRSRAALRALRLRCRTQCAADFPWELDPDWSPSQAQSGAGALRDLSPFGGLIRRAACL 120  
 DB 61 ALRSRAALRALRLRCRTQCAADFPWELDPDWSPSQAQSGAALRDLSPFGGLIRRAACL 120  
 QY 121 RRCLGPPAAHSLSSEMELEFRKSPYNYLOVAYFKINKLEKAVAAAHFFVGNPHEMEMQ 180  
 DB 121 RRCLGPPAAHSLSSEMELEFRKSPYNYLOVAYFKINKLEKAVAAAHFFVGNPHEMEMQ 180  
 QY 181 QNLDYYQTMGSKVEADFDKLETPQHMQRFLGRVRLYSEEQPOEAVPHLEALQYFVAYE 240  
 DB 181 QNLDYYQTMGSKVEADFDKLETPQHMQRFLGRVRLYSEEQPOEAVPHLEALQYFVAYE 240  
 QY 241 ECRALCEGPDYDGYNYLEYNADLFQAITDHYIQLVNLCKQNCVTELASHSPREKPFDFL 300  
 DB 241 ECRALCEGPDYDGYNYLEYNADLFQAITDHYIQLVNLCKQNCVTELASHSPREKPFDFL 300  
 QY 301 PSHYNYLQFAYYNYGNYTQA 320  
 DB 301 PSHYNYLQFAYYNYGNYTQA 320

Search completed: July 18, 2004, 09:58:11  
 Job time : 64 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: July 18, 2004, 09:30:33 ; Search time 87 seconds  
(without alignments)  
2669.211 Million cell updates/sec

Title: US-10-045-815-4  
Perfect score: 3870  
Sequence: 1 MAVRALKLLTLLAVVAAS.....PPEPAQESLSGSESKPKDEL 736

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 1017041 seqs, 315518202 residues

Total number of hits satisfying chosen parameters: 1017041

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : SPTREMBL 25:\*

- 1: sp\_archaea:\*
- 2: sp\_bacteria:\*
- 3: sp\_fungi:\*
- 4: sp\_human:\*
- 5: sp\_invertebrate:\*
- 6: sp\_mammal:\*
- 7: sp\_mnc:\*
- 8: sp\_organelle:\*
- 9: sp\_phase:\*
- 10: sp\_plant:\*
- 11: sp\_rodent:\*
- 12: sp\_virus:\*
- 13: sp\_vertebrate:\*
- 14: sp\_unclassified:\*
- 15: sp\_viruss:\*
- 16: sp\_bacteriap:\*
- 17: sp\_archaeap:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	3870	100.0	736	4 Q9HC86	Q9hc86 homo sapien
2	3846	99.4	736	4 Q96SK8	Q96sk8 homo sapien
3	3844	99.3	736	4 Q96SL5	Q96sl5 homo sapien
4	3582	92.6	804	4 Q96SN3	Q96sn3 homo sapien
5	3403.5	87.9	739	11 Q8CFL3	Q8cfl3 mus musculus
6	3379	87.3	738	11 Q8BSV2	Q8bsv2 mus musculus
7	3378.5	87.3	728	11 Q9RLJ8	Q9rlj8 rattus norv
8	3316	85.7	747	11 Q9QT77	Q9qzt7 mus musculus
9	3305	85.4	742	11 Q9CWK5	Q9cwk5 mus musculus
10	2384	61.6	542	11 Q9QZT6	Q9qzt6 mus musculus
11	2024	52.3	390	4 Q96BR8	Q96br8 homo sapien
12	2014	52.0	390	4 Q9H6K3	Q9h6k3 homo sapien
13	1911	49.4	363	4 Q9HC87	Q9hc87 homo sapien
14	1747	45.1	708	4 Q8IVL5	Q8ivl5 homo sapien
15	1715.5	44.3	703	11 Q8CG71	Q8cg71 mus musculus
16	1473	38.1	736	4 Q8IVL6	Q8ivl6 homo sapien

17	1421.5	36.7	732	11 Q8CG70	Q8cg70 mus musculus
18	1409.5	36.4	527	4 Q9NV12	Q9nvi2 homo sapien
19	1179.5	30.5	551	4 Q13512	Q13512 homo sapien
20	1179.5	30.5	551	4 Q15740	Q15740 homo sapien
21	1127.5	29.1	545	11 Q88836	Q88836 mus musculus
22	1106.5	28.6	409	11 Q8C673	Q8c673 mus musculus
23	672	17.4	412	5 Q9NDP6	Q9ndp6 ciona intes
24	554.5	14.3	388	4 Q8NB01	Q8nb01 homo sapien
25	536.5	13.9	400	11 Q8C8C5	Q8c8c5 mus musculus
26	498	12.9	443	11 Q8K2B0	Q8k2b0 mus musculus
27	474.5	12.3	457	11 Q8C112	Q8c112 mus musculus
28	336	8.7	476	5 Q9NA33	Q9na33 caenorhabdi
29	173.5	4.5	315	11 Q9D1B6	Q9d1b6 mus musculus
30	163	4.2	319	4 Q9H6J2	Q9h6j2 homo sapien
31	144	3.7	383	10 Q9C9X2	Q9c9x2 arabidopsis
32	144	3.7	1025	2 Q84IL2	Q84il2 clostridium
33	141	3.6	342	16 Q67735	Q67735 aquifex aeo
34	135	3.5	1270	4 Q96JN2	Q96jn2 homo sapien
35	132	3.4	197	16 Q8DKV0	Q8dkv0 synecococc
36	129	3.3	227	4 Q13513	Q13513 homo sapien
37	127.5	3.3	1711	17 Q8Q0F8	Q8q0f8 methanosarc
38	126.5	3.2	1075	5 Q8WQ25	Q8wg25 leishmania
39	125.5	3.2	1033	4 Q9UF81	Q9uf81 homo sapien
40	125.5	3.2	1440	3 Q871S3	Q871s3 neurospora
41	123.5	3.2	640	17 Q9V0S2	Q9v0s2 pyrococcus
42	123	3.2	1309	16 Q8CX10	Q8cx10 streptococc
43	123	3.2	1323	16 Q8E5G4	Q8e5g4 streptococc
44	122.5	3.2	1831	10 Q7XN11	Q7xn11 oryza sativ
45	122	3.2	1529	10 Q81068	Q81068 arabidopsis

ALIGNMENTS

RESULT 1

Q9HC86	PRELIMINARY;	PRT;	736 AA.
ID	Q9HC86		
AC	Q9HC86;		
DT	01-MAR-2001 (TREMELrel. 16, Created)		
DT	01-MAR-2001 (TREMELrel. 16, Last sequence update)		
DE	01-OCT-2003 (TREMELrel. 25, Last annotation update)		
DE	GROSL-L protein.		
OS	Homo sapiens (Human)		
OC	Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;		
OC	Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.		
OX	NCBI_TaxID=9606;		
RN	[1]		
RP	SEQUENCE FROM N.A.		
RC	TISSUE=Testis;		
RX	MEDLINE=20406537; PubMed=10951563;		
RA	Kaul S.C., Sugihara T., Yoshida A., Nomura H., Wadhwa R.;		
RT	Grosel, a potential growth suppressor on chromosome 1: its identity to		
RL	baseament membrane-associated, proteoglycan, leprecan."		
RL	Oncogene 19:3576-3583(2000).		
DR	EMBL; AF097432; TAG31019.1;		
DR	GO; GO:0016706; P:oxidoreductase activity, acting on paired d. . . ; IEA.		
DR	GO; GO:0019538; P:protein metabolism; IEA.		
DR	InterPro; IPR005123; 2OG-Fell_Oxy.		
DR	InterPro; IPR000886; ER_target_S.		
DR	InterPro; IPR006620; PR_4_Hyd_alph.		
DR	InterPro; IPR008941; TPR-like.		
DR	Pfam; PF03171; 2OG-Fell_Oxy; 1.		
DR	SMART; SM00702; P4HG; 1.		
DR	PROSITE; PS00014; ER_TARGET; 1.		
SQ	SEQUENCE 736 AA; 83413 MW; 59D1205DB76ADB50 CRC64;		

Query Match 100.0%; Score 3870; DB 4; Length 736;  
Best Local Similarity 100.0%; Pred. No. 1.3e-282;  
Matches 736; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MAVRALKLLTLLAVVAASQAQVESEAGWGVMVTPDLLFAEGTAAYARGDPGVLSMER 60  
DB 1 MAVRALKLLTLLAVVAASQAQVESEAGWGVMVTPDLLFAEGTAAYARGDPGVLSMER 60

QY 61 ALRRAALRALRLRRTCAADFPWELDPDWSPPAQASGAGALRDLSPFGGLLRAACL 120  
 DB 61 ALRRAALRALRLRRTCAADFPWELDPDWSPPAQASGAGALRDLSPFGGLLRAACL 120  
 QY 121 RRLGPPAAHSLSEMELEFRKSPYNYLOVAYFKINKLEKAVAAAHTEFFVGNPEHMEMQ 180  
 DB 121 RRLGPPAAHSLSEMELEFRKSPYNYLOVAYFKINKLEKAVAAAHTEFFVGNPEHMEMQ 180  
 QY 181 QNLDDYQTMGSKVKEADFKDLETPHMQEFLGVRLYSEEQPQEAHPHLEALQEFYFVAYE 240  
 DB 181 QNLDDYQTMGSKVKEADFKDLETPHMQEFLGVRLYSEEQPQEAHPHLEALQEFYFVAYE 240  
 QY 241 ECRALCEGPDYDGVNLYEYNADLFOAITDHYIQVLNCKQNCVTELASHSPREKPFEDFL 300  
 DB 241 ECRALCEGPDYDGVNLYEYNADLFOAITDHYIQVLNCKQNCVTELASHSPREKPFEDFL 300  
 QY 301 PSHYNYLOFAYYNIQNTQAGECAKTYLLFPFNDVNNQNLAYYAAAMLGEHTRSIGPRE 360  
 DB 301 PSHYNYLOFAYYNIQNTQAGECAKTYLLFPFNDVNNQNLAYYAAAMLGEHTRSIGPRE 360  
 QY 361 SAKYRQBSLLEKELLFPAYDVGIPFVDPDSWTPPEVIPKRLQEKOKSERETAVRISQE 420  
 DB 361 SAKYRQBSLLEKELLFPAYDVGIPFVDPDSWTPPEVIPKRLQEKOKSERETAVRISQE 420  
 QY 421 IGNLMKEITLVEEKTRESLDVSRITREGGPLYEGISLTWNSKLLNGYQVMDGVISD 480  
 DB 421 IGNLMKEITLVEEKTRESLDVSRITREGGPLYEGISLTWNSKLLNGYQVMDGVISD 480  
 QY 481 HECQELQRLTNVAATSGDGYRGQTSPTPNEKFGYVTVFKALKLGQSGKVPLOSALHYNN 540  
 DB 481 HECQELQRLTNVAATSGDGYRGQTSPTPNEKFGYVTVFKALKLGQSGKVPLOSALHYNN 540  
 QY 541 VTEKVRIMESYFRDLTPLYSYSHLVCRTAIEBQVQERKDDSHPVHVDNCILNAETLVC 600  
 DB 541 VTEKVRIMESYFRDLTPLYSYSHLVCRTAIEBQVQERKDDSHPVHVDNCILNAETLVC 600  
 QY 601 VKEPPAYTFRDYSAILYNGDFDGNFYFTLDAKTVAEVPQCGRAVGFSSGTENPHG 660  
 DB 601 VKEPPAYTFRDYSAILYNGDFDGNFYFTLDAKTVAEVPQCGRAVGFSSGTENPHG 660  
 QY 661 VKAVTRGQRCALALWFTLDRPHSRDRVQADDLVKMLFSPPEMDLSQEQPLDAQCGPPEP 720  
 DB 661 VKAVTRGQRCALALWFTLDRPHSRDRVQADDLVKMLFSPPEMDLSQEQPLDAQCGPPEP 720  
 QY 721 AQESLSGSESKPKDEL 736  
 DB 721 AQESLSGSESKPKDEL 736  
 RESULT 2  
 Q96SK8 ID Q96SK8 PRELIMINARY; PRT; 736 AA.  
 AC Q96SK8  
 DT 01-DEC-2001 (TrEMBLrel. 19, Created)  
 DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)  
 DE 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)  
 DE Hypothetical protein FLJ14791 (Hypothetical protein  
 DE NT2R2P2005671).  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 OX NCBI\_TaxID=9606;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RA Isogai T., Ota T., Hayashi K., Sugiyama T., Otsuki T., Suzuki Y.,  
 RA Nishikawa T., Nagai K., Sugano S., Aotsuka S., Yoshikawa Y.,  
 RA Matsunawa H., Ishii S., Kawai Y., Saito K., Yamamoto J., Wakamatsu A.,  
 RA Nakamura Y., Nagahari K., Masuho Y., Sasaki N.;  
 RT "NEDO human cDNA sequencing project";  
 RL Submitted (MAY-2001) to the EMBL/GenBank/DBJ databases.  
 RN [2]  
 RP SEQUENCE FROM N.A.

RA Ota T., Nishikawa T., Suzuki Y., Kawai-Hio Y., Hayashi K., Ishii S.,  
 RA Saito K., Yamamoto J., Wakamatsu A., Nagai T., Nakamura Y.,  
 RA Nagahari K., Sugano S., Isogai T.;  
 RT "HRI human cDNA sequencing project";  
 RL Submitted (MAR-2002) to the EMBL/GenBank/DBJ databases.  
 DR EMBL; AK027697; BAB55305.1; -;  
 DR EMBL; AK075418; BAC11608.1; -;  
 DR InterPro; IPR005123; 2OG-Fell\_Oxy.  
 DR InterPro; IPR000886; ER-target\_Oxy.  
 DR InterPro; IPR008941; TPR-like.  
 DR Pfam; PF03171; 2OG-Fell\_Oxy; 1.  
 DR PROSITE; PS00014; ER-TARGET; 1.  
 KW Hypothetical protein\_  
 SQ SEQUENCE 736 AA; 83393 MW; EA1909828FAB685E CRC64;  
 Query Match 99.4%; Score 3846; DB 4; Length 736;  
 Best Local Similarity 99.6%; Pred. NO. 8.6e-281;  
 Matches 733; Conservative 0; Mismatches 3; Indels 0; Gaps 0;  
 QY 1 MAVRALKLLTLLAVVAASQAEVESEAGWMTVDLLFAEGTAAYARGDMPGVLSMER 60  
 DB 1 MAVRALKLLTLLAVVAASQAEVESEAGWMTVDLLFAEGTAAYARGDMPGVLSMER 60  
 QY 61 ALRRAALRALRLRRTCAADFPWELDPDWSPPAQASGAGALRDLSPFGGLLRAACL 120  
 DB 61 ALRRAALRALRLRRTCAADFPWELDPDWSPPAQASGAGALRDLSPFGGLLRAACL 120  
 QY 121 RRLGPPAAHSLSEMELEFRKSPYNYLOVAYFKINKLEKAVAAAHTEFFVGNPEHMEMQ 180  
 DB 121 RRLGPPAAHSLSEMELEFRKSPYNYLOVAYFKINKLEKAVAAAHTEFFVGNPEHMEMQ 180  
 QY 181 QNLDDYQTMGSKVKEADFKDLETPHMQEFLGVRLYSEEQPQEAHPHLEALQEFYFVAYE 240  
 DB 181 QNLDDYQTMGSKVKEADFKDLETPHMQEFLGVRLYSEEQPQEAHPHLEALQEFYFVAYE 240  
 QY 241 ECRALCEGPDYDGVNLYEYNADLFOAITDHYIQVLNCKQNCVTELASHSPREKPFEDFL 300  
 DB 241 ECRALCEGPDYDGVNLYEYNADLFOAITDHYIQVLNCKQNCVTELASHSPREKPFEDFL 300  
 QY 301 PSHYNYLOFAYYNIQNTQAGECAKTYLLFPFNDVNNQNLAYYAAAMLGEHTRSIGPRE 360  
 DB 301 PSHYNYLOFAYYNIQNTQAGECAKTYLLFPFNDVNNQNLAYYAAAMLGEHTRSIGPRE 360  
 QY 361 SAKYRQBSLLEKELLFPAYDVGIPFVDPDSWTPPEVIPKRLQEKOKSERETAVRISQE 420  
 DB 361 SAKYRQBSLLEKELLFPAYDVGIPFVDPDSWTPPEVIPKRLQEKOKSERETAVRISQE 420  
 QY 421 IGNLMKEITLVEEKTRESLDVSRITREGGPLYEGISLTWNSKLLNGYQVMDGVISD 480  
 DB 421 IGNLMKEITLVEEKTRESLDVSRITREGGPLYEGISLTWNSKLLNGYQVMDGVISD 480  
 QY 481 HECQELQRLTNVAATSGDGYRGQTSPTPNEKFGYVTVFKALKLGQSGKVPLOSALHYNN 540  
 DB 481 HECQELQRLTNVAATSGDGYRGQTSPTPNEKFGYVTVFKALKLGQSGKVPLOSALHYNN 540  
 QY 541 VTEKVRIMESYFRDLTPLYSYSHLVCRTAIEBQVQERKDDSHPVHVDNCILNAETLVC 600  
 DB 541 VTEKVRIMESYFRDLTPLYSYSHLVCRTAIEBQVQERKDDSHPVHVDNCILNAETLVC 600  
 QY 601 VKEPPAYTFRDYSAILYNGDFDGNFYFTLDAKTVAEVPQCGRAVGFSSGTENPHG 660  
 DB 601 VKEPPAYTFRDYSAILYNGDFDGNFYFTLDAKTVAEVPQCGRAVGFSSGTENPHG 660  
 QY 661 VKAVTRGQRCALALWFTLDRPHSRDRVQADDLVKMLFSPPEMDLSQEQPLDAQCGPPEP 720  
 DB 661 VKAVTRGQRCALALWFTLDRPHSRDRVQADDLVKMLFSPPEMDLSQEQPLDAQCGPPEP 720  
 QY 721 AQESLSGSESKPKDEL 736  
 DB 721 AQESLSGSESKPKDEL 736

Db	541	VTEKARIMESFRLDTPIYFSYSHLVCRITAEVQAEKDSHPVHVDNCILNAETLVC	600
Qy	601	VKEPPATYTRDYSAILLYLNGDFDGGNFYFTELDPAKTTVAEVQPCGRAVFSFGSTENPHG	660
Db	601	VKEPPATYTRDYSAILLYLNGDFDGGNFYFTELDPAKTTVAEVQPCGRAVFSFGSTENPHG	660
Qy	661	VKAVTGQSCATALWETLDPHSERDVOADDLVKMLFSPPEMDLSQEQPLDAQGPPPEP	720
Db	661	VKAVTGQSCATALWETLDPHSERDVOADDLVKMLFSPPEMDLSQEQPLDAQGPPPEP	720
Qy	721	AQESLSGSESXPKDEL 736	
Db	721	AQESLSGSESXPKDEL 736	
RESULT 4			
ID	Q96SN3	PRELIMINARY; PRT; 804 AA.	
AC	Q96SN3;		
DT	01-DEC-2001 (T-EMBLrel. 19, Created)		
DT	01-DEC-2001 (T-EMBLrel. 19, Last sequence update)		
DT	01-OCT-2003 (T-EMBLrel. 25, Last annotation update)		
DE	Hypothetical protein FLJ14742.		
OS	Homo sapiens (human).		
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;		
OC	Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.		
OC	NCBI_TaxID:9606;		
RN	[1]		
RP	SEQUENCE FROM N.A.		
RA	Isegai T., Ota T., Hayashi K., Sugiyama T., Otsuki T., Suzuki Y.,		
RA	Nishikawa T., Nagai K., Sugano S., Shiratori A., Sudo H.,		
RA	Wagatsuma M., Hosoi T., Kaku Y., Kodaira H., Kondo H., Sugawara M.,		
RA	Takahashi M., Chiba Y., Ishida S., Murakawa K., Ono Y., Takiguchi S.,		
RA	Watanabe S., Kimura K., Murakami K., Ishii S., Kawai Y., Saito K.,		
RA	Yanamoto J., Wakamatsu A., Nakamura Y., Nagahari K., Masuho Y.,		
RA	Ninomiya K., Iwayanagi T.;		
RT	"NEDO human cDNA sequencing project.";		
RL	Submitted (MAY-2001) to the EMBL/GenBank/DBJ databases.		
DR	EMBL; AK027648; BAB53264.1; -		
DR	Genbank; HGNC:19316; LEPRE1.		
DR	GO; GO:0016706; F:oxidoreductase activity, acting on paired d. . .; IEA.		
DR	GO; GO:0019338; P:protein metabolism; IEA.		
DR	InterPro; IPR005123; 2OG-Feril_Oxy.		
DR	InterPro; IPR006620; Pro.4_hyd_alph.		
DR	InterPro; IPR008941; TPR-like.		
DR	Pfam; PF03171; 2OG-Feril_Oxy; 1.		
DR	SMART; SM00702; P4HC; 1.		
KW	Hypothetical protein.		
SQ	SEQUENCE 804 AA; 90544 MW; 23C8D9E503D208CD CRC64;		
Query Match			
Best Local Similarity 99.3%; Score 3582; DB 4; Length 804;			
Matches 682; Conservative 0; Mismatches 5; Indels 0; Gaps 0;			
Qy	1	MAYRAKLLITLLAVVAASQAEVESAGWMTPTDLLFAEGTAAYARGDWPGVVLNWER	60
Db	1	MAYRAKLLITLLAVVAASQAEVESAGWMTPTDLLFAEGTAAYARGDWPGVVLNWER	60
Qy	61	ALRSRAALRALRLRCRTQCAADFPWELDPDWSPPSPAQASGAGALRDLSPFGGLLRRACL	120
Db	61	ALRSRAALRALRLRCRTQCAADFPWELDPDWSPPSPAQASGAGALRDLSPFGGLLRRACL	120
Qy	121	RRLCLGPPAAHSLSEEMLEFRKSPNYLVAFYKINKLEKAVAAATFFVGNPEHMEMQ	180
Db	121	RRLCLGPPAAHSLSEEMLEFRKSPNYLVAFYKINKLEKAVAAATFFVGNPEHMEMQ	180
Qy	181	QNLDYYQTMISGVKAEADFKLETQPHMQEFLRGLVLYSEEQPQEAAPVHLEALQEFVAYE	240
Db	181	QNLDYYQTMISGVKAEADFKLETQPHMQEFLRGLVLYSEEQPQEAAPVHLEALQEFVAYE	240
Qy	241	ECRALCGPVDYDGYNYLEVNADLFOAITHYIQVLNCKQNCVTELASHSREKPFDFDL	300

```

Db 241 ECRCALCEGYDYGNYLYNADLFQAITDHYIQLVNCNKCVCVTELASHPSREKPFEDFL 300
Qy 301 PSHVNYLQFAYNIGNYTOAGCACTYLLFFPNDEVNQNLAAYAAAMLGEEHTRISIGPRE 360
Db 301 PSHVNYLQFAYNIGNYTOAGCACTYLLFFPNDEVNQNLAAYAAAMLGEEHTRISIGPRE 360
Qy 361 SAKERYORSLEKELELLFFAYDVGIFPFVDPSWTPPEEVIKGLQEKOKSERETAVRISQE 420
Db 361 SAKERYORSLEKELELLFFAYDVGIFPFVDPSWTPPEEVIKGLQEKOKSERETAVRISQE 420
Qy 421 IGNLMKEITLVEEKTESLVSRLTREGGPLYEGISLTNMSKLLNGYQRVVMGDVISD 480
Db 421 IGNLMKEITLVEEKTESLVSRLTREGGPLYEGISLTNMSKLLNGYQRVVMGDVISD 480
Qy 481 HECQELQRLTNVAATSGDGYRGOTSPHTPNEKFGYVTVFKALKLQEGKVPLOSAHLIYN 540
Db 481 HECQELQRLTNVAATSGDGYRGOTSPHTPNEKFGYVTVFKALKLQEGKVPLOSAHLIYN 540
Qy 541 VTEKVRIMESYFRDLTPLYFSYSHLVCTAIEVQAEKDDSHPVHVDNCILNAETLVC 600
Db 541 VTEKVRIMESYFRDLTPLYFSYSHLVCTAIEVQAEKDDSHPVHVDNCILNAETLVC 600
Qy 601 VKEPPAYTFRDYSAILYNGDFDGNFYFTLDAKTVAEVQPCGRAVGFSSGTENPHG 660
Db 601 VKEPPAYTFRDYSAILYNGDFDGNFYFTLDAKTVAEVQPCGRAVGFSSGTENPHG 660
Qy 661 VKAVTRGORCAIALWFTLDRHSEDR 687
Db 661 VKAVTRGORCAIALWFTLDRHSEDR 687

RESULT 5
Q8CFL3 PRELIMINARY; PRT; 739 AA.
AC 08CFL3
DT 01-MAR-2003 (TrEMBLrel. 23, Created)
DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Similar to leprecan.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
CX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J; TISSUE=Breast tumor;
RA Strausberg R.;
RL Submitted (FEB-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL; BC024047; AAH24047.1; -.
DR PIR; P00551; PT0551.
DR GO; GO:0016706; F:oxidoreductase activity, acting on paired d. . .; IEA.
DR GO; GO:0019538; P:protein metabolism; IEA.
DR InterPro; IPR005123; 2OG-FeII_Oxy.
DR InterPro; IPR008886; ER target S.
DR InterPro; IPR006620; Pro_4_hyd_alph.
DR InterPro; IPR008941; TPR-like.
DR Pfam; PF03171; 2OG-FeII_Oxy; 1.
DR SMART; SM00702; F4Hc; 1.
DR PROSITE; PS00014; ER_TARGET; 1.
SQ SEQUENCE 739 AA; 83592 MW; 64B7AE68E80B68EA CRC64;

Query Match 87.9%; Score 3403.5; DB 11; Length 739;
Best Local Similarity 87.4%; Pred No. 1.9e-247;
Matches 646; Conservative 34; Mismatches 56; Indels 3; Gaps 2;

Qy 1 MAVRALKLLTLLAVAAAS-QAEVESEAGGMVTPDLLFAEGTAAYAGDWPGLVLSME 59
Db 1 MAVSERLLAAMLAIAAAALRAVAASEPGWDVAAPDLLYAEGTAAYAGDWPGLVLSME 60
Qy 60 RALSRALRALRLRCCTQCAADPPWELDPDWSPP--ACASGAGALRDLSPFGGLRRA 117
Db 61 RALSRALRALRLRCCTQCAADPPWELDPDWSPP--ACASGAGALRDLSPFGGLRRA 120

```

```

Qy 118 ACLRCLGPPAAHSISEMELFEFRKSPYNYLOVAYFKINKLEKAVAAAHFTFVGNPEHM 177
Db 121 ACLRCLGPPSAHLSSELDLEFNKRSYPNYLOVAYFKINKLEKAVAAAHFTFVGNPEHM 180
Qy 178 EMQNLDYYQTMSGVKEADFQDLETPQHMQEPRGLVRLYSEBQPOEAVPHLEAALQEVFV 237
Db 181 EMQNLDYYQTMSGVKEADFQDLEAKPHMHEPRGLVRLYSEBQPOEAVPHLEAALQEVFV 240
Qy 238 AVECRALCEGYDYGNYLYNADLFQAITDHYIQLVNCNKCVCVTELASHPSREKPF 297
Db 241 ADEECRALCEGYDYGNYLYNADLFQAITDHYVQVQVNCNKCVCVTELASHPSREKPF 300
Qy 298 DFLPHVNYLQFAYNIGNYTOAGCACTYLLFFPNDEVNQNLAAYAAAMLGEEHTRISIG 357
Db 301 DFLPHVNYLQFAYNIGNYTOAIECACTYLLFFPNDEVNQNLAAYTAMLGEAEASIS 360
Qy 358 PRESAREYRQSRLLLEKELLFFAYDVGIFPFVDPSWTPPEEVIKGLQEKOKSERETAVRI 417
Db 361 PRENAEYRRLSLEKELLFFAYDVGIFPFVDPSWTPPEEVIKGLQEKOKSERETAVRI 420
Qy 418 SOEIGNLMKEITLVEEKTESLVSRLTREGGPLYEGISLTNMSKLLNGYQRVVMGDV 477
Db 421 SOEIGNLMKEITLVEEKTESLVSRLTREGGPLYEGISLTNMSKLLNGYQRVVMGDV 480
Qy 478 ISDHCEQELQRLTNVAATSGDGYRGOTSPHTPNEKFGYVTVFKALKLQEGKVPLOSAHL 537
Db 481 ISDDECEQELQRLTNVAATSGDGYRGOTSPHTPNEKFGYVTVFKALKLQEGKVPLOSAHM 540
Qy 538 YNNVTEKVRIMESYFRDLTPLYFSYSHLVCTAIEVQAEKDDSHPVHVDNCILNAET 597
Db 541 YNNVTEKVRIMESYFRDLTPLYFSYSHLVCTAIEESQAEKDDSHPVHVDNCILNAEA 600
Qy 598 LVCVKEPPAYTFRDYSAILYNGDFDGNFYFTLDAKTVAEVQPCGRAVGFSSGTEN 657
Db 601 LNCIKEPPAYTFRDYSAILYNGDFDGNFYFTLDAKTVAEVQPCGRAVGFSSGTEN 660
Qy 658 PHGVKAVTRGORCAIALWFTLDRHSEDRVQADDLVKMLSPPEMDLSQBPDLDAQGP 717
Db 661 PHGVKAVTRGORCAIALWFTLDRHSEDRVQADDLVKMLSPPEMDLSQBPDLDAQGP 720
Qy 718 PEPAOESSLSGSESKEPKDEL 736
Db 721 PEFGEESLSDRGSLHKDEL 739

RESULT 6
Q8BSV2 PRELIMINARY; PRT; 738 AA.
AC Q8BSV2
DT 01-MAR-2003 (TrEMBLrel. 23, Created)
DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Leprecaan 1.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
CX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J; TISSUE=Pituitary;
RX MEDLINE=22354683; PubMed=12466851;
RA The FANTOM Consortium,
RT the RIKEN Genome Exploration Research Group Phase I & II Team;
RT "Analysis of the mouse transcriptome based on functional annotation of
60,770 full-length cDNAs.";
RL Nature 420:563-573 (2002).
DR EMBL; AK030436; BAC26962.1; -.
DR PIR; PT0551; PT0551.
DR GO; GO:0016706; F:oxidoreductase activity, acting on paired d. . .; IEA.
DR GO; GO:0019538; P:protein metabolism; IEA.
DR InterPro; IPR005123; 2OG-FeII_Oxy.
DR InterPro; IPR008886; ER_target_S.
DR InterPro; IPR006620; Pro_4_hyd_alph.

```

Query Match	87.3%	Score 3379	DB 11	Length 738
Best Local Similarity	87.1%	Pred. No. 1.3e-245		
Matches 644	Conservative 34	Mismatches 57	Indels 4	Gaps 3
DR InterPro: IPR008941; TPR-like.				
DR Pfam: PF03171; 2OG-Ferri Oxy; 1.				
DR SMART: SM00702; P4HC; 1.				
DR PROSITE: PS00014; ER TARGET; 1.				
SQ SEQUENCE 738 AA; 83528 MW; 75967DE318F55F4A CRC64;				
QY 1 MAVRALKLLTLLAVAAAS-QAEVSESGAGMGTVPDLLFAEGTAAYARGDGPGLVLSME 59				
DB 1 MAVERRLLAAMLAVAAAAALRVAASEPGWDVAAPDLLLYAEGTAAYSRGDGPGLVLSME 60				
QY 60 RALRSRAALRALRLRCRTQCAADTPWELDPWSP--AQASGAGALRDLSPFFGGLRRA 117				
DB 61 RALRSRAALRALRLRCRTCAATELFWAPDLDLGDPDPSLSDPGGAALHDLFFGAVLRA 120				
QY 118 ACLRCLGPPAAHLSSEMELEFRKRSYNYLQVAYFKINKLEKAVAAAHFFVGVNPEHM 177				
DB 121 ACLRCLGPPSAHLLSELDLEFRKRSYNYLQVAYFKINKLEKAVAAAHFFVGVNPEHM 180				
QY 178 EMQNLDYYQTMGSGVKEADFQDLETPHQMQEFRIGLVRLYSEEQPQEAAPHLEALQBYFV 237				
DB 181 EMQNLDYYQTMGSGVKEADFRLDAKPHMHSEFRIGLVRLYSEKQEAAPHLEALQBYFV 240				
QY 238 AVEECRALCEGPDYDGYNYLEYNAIDLFOALTIDHYIOVLNCKQNCVTELASHPSREKPF 297				
DB 241 ADEECRALCEGPDYDGYNYLDYSDLEFQALTIDHYIOVLNCKQNCVTELASHPSREKPF 300				
QY 298 DFLPSHNYLQFAYYINIGNYTAQGECAKTYLLFFPNDENVNQNLAYYAAMLGEBHRSIG 357				
DB 301 DFLPSHNYLQFAYYINIGNYTAQTECAKTYLLFFPNDENVHQNLAAYTAMLGEDEASSIS 360				
QY 358 PRESAGEYRQSRLLLEKELLFPAYDVFGIPFVDPDPSWTPEEVIKLEQEKQKSERETAARI 417				
DB 361 PRENAEYRRRSLLLEKELLFPAYDIFGIPFVDPDPSWTPEEVIKRLQEKQKSERETAAR- 419				
QY 418 SOETGNLMKETETLVEEKTESLDVSRITREGGPLLIEGISLTWNKSLINGYQRVWMDGV 477				
DB 420 SOETGNLMKETETLVEEKTESLDVSRITREGGPLLIEGISLTWNKSVLNGSRVWMDGV 479				
QY 478 ISDHECQELQRLTNVAATSGDGYRGQTSPTPTNEKPYGVTVFKALKLQEGKVPLOSAHL 537				
DB 480 ISDDECQELQRLTNAAAATSGDGYRGQTSPTPTNEKPYGVTVFKALKLQEGKVPLOSAHM 539				
QY 538 YYNVTEKVRRTMESYFRDLTPLYTSVSHLVCRTALEEVQAEKKDDSPHVHVDNCLINAEI 597				
DB 540 YYNVTEKVRRTMESYFRDLTPLYTSVSHLVCRTALEESQAEKKDDSPHVHVDNCLINAEA 599				
QY 598 LVCVKEPPAYTFRDYSAILYLNGDDPDGNGFYFTELDKATVTAEVQPCQRAVGFSSGTE 657				
DB 600 LMCVKEPPAYTFRDYSAILYLNGDDPDGNGFYFTELDKATVTAEVQPCQRAVGFSSGTE 659				
QY 658 PHGVKAVTRGORCAIALNFTLDPHSEDRDQADDLVVKMLSPSEEMDLISQEQPLDAQOQP 717				
DB 660 PHGVKAVTRGORCAIALNFTLDPHSEDRDQADDLVVKMLSPSEEVLDLPQEQPLDAQOQS 719				
QY 718 PEPAQESLSGSGESKPKDEL 736				
DB 720 PEPGEESLSDRGSLHKDEL 738				
RESULT 7				
ID Q9RLJ8 PRELIMINARY; PRT: 728 AA.				
AC Q9RLJ8;				
DT 01-MAY-2000 (TREMELrel. 13, Created)				
DT 01-MAY-2000 (TREMELrel. 13, Last sequence update)				
DE 01-OCT-2003 (TREMELrel. 25, Last annotation update)				
DE Leprecan.				
OS Rattus norvegicus (Rat).				

PRECISE. T 7

RESULI 7	Q9RLJ8	PRELIMINARY;	PRT;	728 AA.
ID	Q9RLJ8			
AC	Q9RLJ8;			
DT	01-MAY-2000	(TReMBLrel. 13, Created)		
DT	01-MAY-2000	(TReMBLrel. 13, Last sequence update)		
DT	01-OCT-2003	(TReMBLrel. 25, Last annotation update)		
DE	Leprecan.			
GN	LEPREL			
OS	Rattus norvegicus (Rat).			



```

QY 729 ESKPKDEL 736
DB 721 ESQPKDEL 728

RESULT 8
Q9QZT7 PRELIMINARY; PRT; 747 AA.
AC Q9QZT7;
DT 01-MAY-2000 (TRENBLrel. 13, Created)
DT 01-MAY-2000 (TRENBLrel. 13, Last sequence update)
DT 01-OCT-2003 (TRENBLrel. 25, Last annotation update)
DE Growth suppressor 1L.
GN LEPRE1 OR GROS1.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=CD1 ICR; TISSUE=Fibroblast;
RX MEDLINE=20406537; PubMed=10951563;
RA Kaul S.C., Sugihara T., Yoshida A., Nomura H., Wadhwa R.;
RT "Gros1, a potential growth suppressor on chromosome 1: its identity to
RL basement membrane-associated proteoglycan, leprecan.";
DR ENCL; AF165163; AAF04806.1; -.
DR MGD; MGI:1888921; Leprel.
DR GO; GO:0005737; C:cytoplasm; IDA.
DR GO; GO:0005634; C:nucleus; IDA.
DR GO; GO:0016049; P:cell growth; IDA.
DR InterPro; IPR005123; 2OG-FeII_Oxy.
DR InterPro; IPR008941; TPR-like.
DR Pfam; PF03171; 2OG-FeII_Oxy; 1.
DR SMART; SM00702; P4Hc; 1.
SQ SEQUENCE 747 AA; 84811 MW; EBEAB2A62E824123 CRC64;

Query Match 85.7%; Score 3316; DB 11; Length 747;
Best Local Similarity 88.2%; Pred. No. 7.4e-241;
Matches 623; Conservative 33; Mismatches 48; Indels 2; Gaps 1;

QY 25 ESEAGWGWTPDLFARGTAAAYAGDWPVGVLSMERALRSALRALRLRCRTQCAADFP 84
DB 25 ESEFGWVAAPDLLAYAGTAAYSRRDWPVGVLSMERALRSALRALRLRCRTQCAETLP 84
QY 85 WELDPDWSPP--AAQSGAGALRDLSPFGGLLRRAACLRCLGPPAAHLSISEMELEPRK 142
DB 85 WAPDLGLGPPSLSQDPCAAALHDLRFFGAVLRRAACLRCLGPPSAHLLSELDLEFNK 144
QY 143 RSPNYLQVAYEKINKLEKAVAAAHFFVGNPHEMQRQNLQVYQTMGSKVKEADPKDLET 202
DB 145 RSPNYLQVAYEKINKLEKAVAAAHFFVGNPHEMQRQNLQVYQTMGSKVKEADPKDLEA 204
QY 203 QPHMQEPLGLVRLYSEBPQEAIPHLEAALQEFVAYEECRALCEGPDYDGVNLYEYNA 262
DB 205 KPHMHEPLGLVRLYSEBKPQEAIPHLEAALQEFVAYEECRALCEGPDYDGVNLYDYSA 264
QY 263 DLQQAITDHYIOLVNLCKQNCVTSLASHPSSEKPFEDPLSHYNYLQFAYNYIGNYTOAGE 322
DB 265 DLQQAITDHYVQVNLCKQNCVTSLASHPSSEKPFEDPLSHYNYLQFAYNYIGNYTOAIE 324
QY 323 CAKTYLLFFPNDVNMQLAYYAAMLGEEHTRSIGPREGSAKEYRQSRSLLEKELLFPAYDV 382
DB 325 CAKTYLLFFPNDVNMQLAYYAAMLGEEHTRSIGPREGSAKEYRQSRSLLEKELLFPAYDI 384
QY 383 FGIPFPVDDSWTPEBVIKRLQSKQKSERETAVRISQEIQLNLMKEITLVEEKTKESLDV 442
DB 385 FGIPFPVDDSWTPEBVIKRLQSKQKSERETAVRISQEIQLNLMKEITLVEEKTKESLDV 444
QY 443 SRLTREGGPLYEGISLTWNSKVLNGSRQVVMGDI SDHECOBLQRLTNVAATSGDGYRG 502
DB 445 SRLTREGGPLYEGISLTWNSKVLNGSRQVVMGDI SDDECEQLQRLTNAAATSGDGYRG 504
QY 503 QTSPTNEKEFYGVTVFKALKLGQEGKVPLOSALHYLYNNVTEKTRIMESYRLDTPLYFS 562
DB 505 QTSPTNEKEFYGVTVFKALKLGQEGKVPLOSALHYLYNNVTEKTRIMESYRLDTPLYFS 564
QY 563 YSHLVCRTAIEBVOAERKDDSHPHVDNCILNAETLVCVKEPPAYTRFDYSAILYLNQDF 622
DB 565 YSHFVCRTAIEBSQAERKDDSHPHVDNCILNAEAFMCIEPPAYTPREYSAILYLNQDF 624
QY 623 DCGNFYFTELDKATVTAEVQPCGSAVGFSGGTENPHGVKAVTRGRCORCAIALWTLDPRH 682
DB 625 DCGNFYFTELDKATVTAEVQPCGSAVGFSGGTENPHGVKAVTRGRCORCAIALWTLDPRH 684
QY 683 SERDRVQADDLVKMLFSPPEMDLSQEQPLDAQQPPPEPAQESLSGS 728
DB 685 SERDRVQADDLVKMLFSPPEMDLSQEQPLDAQQPPPEPAQESLSGS 730

RESULT 9
Q9CWK5 PRELIMINARY; PRT; 742 AA.
AC Q9CWK5;
DT 01-JUN-2001 (TRENBLrel. 17, Created)
DT 01-JUN-2001 (TRENBLrel. 17, Last sequence update)
DT 01-OCT-2003 (TRENBLrel. 25, Last annotation update)
DE 2410024C15Rik protein.
GN LEPRE1 OR GROS1 OR 2410024C15RIK.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J; TISSUE=Embryonic stem cells;
RX MEDLINE=21085660; PubMed=11217851;
RA Kawai J., Shinagawa A., Shibata K., Yoshino M., Itoh M., Iehi Y.,
RA Arakawa T., Hara A., Fukunishi Y., Konno H., Adachi J., Fukuda S.,
RA Aizawa K., Izawa M., Nishi K., Kiyosawa H., Kondo S., Yamanaoka I.,
RA Saito T., Okazaki Y., Gojobori T., Bono H., Kasukawa T., Saito R.,
RA Kadota K., Matsuda H.A., Ashburner M., Batalov S., Casavant T.,
RA Fleischmann W., Gaasterland T., Gissi C., King B., Kochiwa H.,
RA Kuehl P., Lewis S., Matsuo Y., Nikaido I., Pesole G., Quackenbush J.,
RA Schriml L.M., Staubli F., Suzuki R., Tomita M., Wagner L., Washio T.,
RA Sakai K., Okido T., Furuno M., Aono H., Baldarelli R., Barsh G.,
RA Blake J., Boffelli D., Bojunga N., Carninci P., de Bonaldo M.F.,
RA Brownstein M.J., Bult C., Fletcher C., Fujita M., Gariboldi M.,
RA Gustincich S., Hill D., Hofmann M., Hume D.A., Kamiya M., Lee N.H.,
RA Lyons P., Marchionni L., Mashima J., Mazzarelli J., Mombaerts P.,
RA Nordone P., Ring B., Ringwald M., Rodriguez I., Sakamoto N.,
RA Sasaki H., Sato K., Schoenbach C., Seya T., Shibata Y., Storch K.-F.,
RA Suzuki H., Toyooka K., Wang X.H., Weitz C., Whittaker C., Wilming L.,
RA Wyshaw-Boris A., Yoshida K., Hasegawa Y., Kawaji H., Kotsuki S.,
RA Hayashizaki Y.;
RT "Functional annotation of a full-length mouse cDNA collection.";
RL Nature 409:685-690(2001).
DR EMBL; AK010578; BAB27041.1; -.
DR MGD; MGI:1888921; Leprel.
DR GO; GO:0005737; C:cytoplasm; IDA.
DR GO; GO:0005634; C:nucleus; IDA.
DR GO; GO:0016049; P:cell growth; IDA.
DR InterPro; IPR005123; 2OG-FeII_Oxy.
DR InterPro; IPR008941; TPR-like.
DR Pfam; PF03171; 2OG-FeII_Oxy; 1.
DR SMART; SM00702; P4Hc; 1.
SQ SEQUENCE 742 AA; 84086 MW; CFC44D34D96ACEA7 CRC64;

Query Match 85.4%; Score 3305; DB 11; Length 742;
Best Local Similarity 86.0%; Pred. No. 4.9e-240;
Matches 629; Conservative 37; Mismatches 57; Indels 8; Gaps 3;

QY 1 MAVZALKLLTLLAVAAAS-QAEVSEAGWGVTPDLLFAEGTAAVARGDWPVLSME 59

```





RT vector.";  
 RL Submitted (MAY-2003) to the EMBL/GenBank/DBJ databases.  
 DR EMBL; BC015309; AAH15309.1; -.  
 DR EMBL; BT007039; AAP35688.1; -.  
 DR GO; GO:0016706; P:oxidoreductase activity, acting on paired d. . .; IEA.  
 DR GO; GO:0019538; P:protein metabolism; IEA.  
 DR InterPro; IPR005123; 2OG-Fell\_Oxy.  
 DR InterPro; IPR000886; ER\_target\_S.  
 DR InterPro; IPR006620; Pro\_4\_hyd\_alph.  
 DR Pfam; PF03171; 2OG-Fell\_Oxy; 1.  
 DR SMART; SM00702; P4HC; 1.  
 DR PROSITE; PS00014; ER\_TARGET; 1.  
 DR SQ SEQUENCE 390 AA; 44077 MW; 61C9F5983CF637FB CRC64;  
 Query Match 52.3%; Score 2024; DB 4; Length 390;  
 Best Local Similarity 99.7%; Pred. No. 5.7e-144;  
 Matches 389; Conservative 0; Mismatches 1; Indels 0; Gaps 0;  
 QY 347 MLGEHTRISIGPRESAKYRQSLLEKELFFAYDVFGIPFVDPDPSWTPEEVIKRLQEK 406  
 DB 1 MLGEHTRISIGPRESAKYRQSLLEKELFFAYDVFGIPFVDPDPSWTPEEVIKRLQEK 60  
 QY 407 QKSERETAVRISQEIQNLNLMKEIETLVVEKTKESLDVSRLTREGGPLYEGISLTMSKLL 466  
 DB 61 QKSERETAVRISQEIQNLNLMKEIETLVVEKTKESLDVSRLTREGGPLYEGISLTMSKLL 120  
 QY 467 NGYQVRVMDGVISDHECOELQRLTNVAATSGDGYRGQTSPTNEKFGYGVTVFKALKLQ 526  
 DB 121 NSQVRVMDGVISDHECOELQRLTNVAATSGDGYRGQTSPTNEKFGYGVTVFKALKLQ 180  
 QY 527 EGKVPLOSAHLIYNNVTEKVRIMESYFRDLTPLYFSYSHLVCRTAIEEVOAERKDDSHPV 586  
 DB 181 EGKVPLOSAHLIYNNVTEKVRIMESYFRDLTPLYFSYSHLVCRTAIEEVOAERKDDSHPV 240  
 QY 587 HVDNCILNAETLVCKEPPATFRDYSAIYLNGDFDGGNFYFTELDAAKTVTAEVQPOCG 646  
 DB 241 HVDNCILNAETLVCKEPPATFRDYSAIYLNGDFDGGNFYFTELDAAKTVTAEVQPOCG 300  
 QY 647 RAVGFSSGTENPHGVKAVTRGQRCALWFTLDPHRSERDRVQADDLVKMLFSPPEMDLS 706  
 DB 301 RAVGFSSGTENPHGVKAVTRGQRCALWFTLDPHRSERDRVQADDLVKMLFSPPEMDLS 360  
 QY 707 QEQLDAQQGPPEPAQESLSGSESKPKDEL 736  
 DB 361 QEQLDAQQGPPEPAQESLSGSESKPKDEL 390  
 RESULT 12  
 Q9H6K3 PRELIMINARY; PRT; 390 AA.  
 AC Q9H6K3  
 DT 01-MAR-2001 (TrEMBLrel. 16, Created)  
 DT 01-MAR-2001 (TrEMBLrel. 16, Last sequence update)  
 DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)  
 DE Hypothetical protein FLJ22188.  
 OS Homo sapiens (Human)  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 OX NCBI\_TaxID=9606;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RA Watanabe K., Kumagai A., Itakura S., Yamazaki M., Tashiro H., Ota T.,  
 RA Suzuki Y., Ohtsuki M., Nishi T., Shibahara T., Tanaka T.,  
 RA Nakamura Y., Isoigaki T., Sugano S.;  
 RT "NEDO human cDNA sequencing project";  
 RL Submitted (AUG-2000) to the EMBL/GenBank/DBJ databases.  
 DR EMBL; AK025841; BAB15256.1; -.  
 DR GO; GO:0016706; F:oxidoreductase activity, acting on paired d. . .; IEA.  
 DR GO; GO:0019538; P:protein metabolism; IEA.  
 DR InterPro; IPR005123; 2OG-Fell\_Oxy.  
 DR InterPro; IPR000886; ER\_target\_S.  
 DR InterPro; IPR006620; Pro\_4\_hyd\_alph.  
 DR Pfam; PF03171; 2OG-Fell\_Oxy; 1.  
 Query Match 49.4%; Score 1911; DB 4; Length 363;  
 Best Local Similarity 100.0%; Pred. No. 1.7e-135;  
 Matches 360; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 NAVFALKLLTLLAVAAASQAEVSEAGAGMTPTDLLFAEGTAAAYARGDWPGLVLSMER 60  
 DB 1 NAVFALKLLTLLAVAAASQAEVSEAGAGMTPTDLLFAEGTAAAYARGDWPGLVLSMER 60  
 QY 61 ALRGAALRALRLCRCTCCAADFFWELDPDWSPPAQAGALRDLSPFGGLLRRAACL 120

DR SMART; SM00702; P4HC; 1.  
 DR PROSITE; PS00014; ER\_TARGET; 1.  
 DR KW Hypothetical protein\_  
 DR SQ SEQUENCE 390 AA; 44093 MW; 61D2999310E677FB CRC64;  
 Query Match 52.0%; Score 2014; DB 4; Length 390;  
 Best Local Similarity 99.5%; Pred. No. 3.2e-143;  
 Matches 388; Conservative 0; Mismatches 2; Indels 0; Gaps 0;  
 QY 347 MLGEHTRISIGPRESAKYRQSLLEKELFFAYDVFGIPFVDPDPSWTPEEVIKRLQEK 406  
 DB 1 MLGEHTRISIGPRESAKYRQSLLEKELFFAYDVFGIPFVDPDPSWTPEEVIKRLQEK 60  
 QY 407 QKSERETAVRISQEIQNLNLMKEIETLVVEKTKESLDVSRLTREGGPLYEGISLTMSKLL 466  
 DB 61 QKSERETAVRISQEIQNLNLMKEIETLVVEKTKESLDVSRLTREGGPLYEGISLTMSKLL 120  
 QY 467 NGYQVRVMDGVISDHECOELQRLTNVAATSGDGYRGQTSPTNEKFGYGVTVFKALKLQ 526  
 DB 121 NSQVRVMDGVISDHECOELQRLTNVAATSGDGYRGQTSPTNEKFGYGVTVFKALKLQ 180  
 QY 527 EGKVPLOSAHLIYNNVTEKVRIMESYFRDLTPLYFSYSHLVCRTAIEEVOAERKDDSHPV 586  
 DB 181 EGKVPLOSAHLIYNNVTEKVRIMESYFRDLTPLYFSYSHLVCRTAIEEVOAERKDDSHPV 240  
 QY 587 HVDNCILNAETLVCKEPPATFRDYSAIYLNGDFDGGNFYFTELDAAKTVTAEVQPOCG 646  
 DB 241 HVDNCILNAETLVCKEPPATFRDYSAIYLNGDFDGGNFYFTELDAAKTVTAEVQPOCG 300  
 QY 647 RAVGFSSGTENPHGVKAVTRGQRCALWFTLDPHRSERDRVQADDLVKMLFSPPEMDLS 706  
 DB 301 RAVGFSSGTENPHGVKAVTRGQRCALWFTLDPHRSERDRVQADDLVKMLFSPPEMDLS 360  
 QY 707 QEQLDAQQGPPEPAQESLSGSESKPKDEL 736  
 DB 361 QEQLDAQQGPPEPAQESLSGSESKPKDEL 390  
 RESULT 13  
 Q9HC87 PRELIMINARY; PRT; 363 AA.  
 AC Q9HC87  
 DT 01-MAR-2001 (TrEMBLrel. 16, Created)  
 DT 01-MAR-2001 (TrEMBLrel. 16, Last sequence update)  
 DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)  
 DE GROS1-S protein.  
 OS Homo sapiens (Human)  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 OX NCBI\_TaxID=9606;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RA Kaul S.C., Sugihara T., Yoshida A., Nomura H., Wadhwa R.;  
 RT "Gros1, a potential growth suppressor on chromosome 1: its identity to  
 RT basement membrane-associated proteoglycan, leprecan.";  
 RL Oncogene 19:3576-3583 (2000).  
 DR EMBL; AF097431; AAG31018.1; -.  
 DR GO; GO:0008285; P:negative regulation of cell proliferation; NAS.  
 DR InterPro; IPR008941; TPR-like.  
 DR SQ SEQUENCE 363 AA; 41158 MW; 5B1B66B38679E76A CRC64;  
 Query Match 49.4%; Score 1911; DB 4; Length 363;  
 Best Local Similarity 100.0%; Pred. No. 1.7e-135;  
 Matches 360; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 NAVFALKLLTLLAVAAASQAEVSEAGAGMTPTDLLFAEGTAAAYARGDWPGLVLSMER 60  
 DB 1 NAVFALKLLTLLAVAAASQAEVSEAGAGMTPTDLLFAEGTAAAYARGDWPGLVLSMER 60  
 QY 61 ALRGAALRALRLCRCTCCAADFFWELDPDWSPPAQAGALRDLSPFGGLLRRAACL 120





GenCore version 5.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - nucleic search, using frame\_plus\_p2n model

Run on: July 18, 2004, 10:02:10 ; Search time 1081 Seconds  
(without alignments)  
3321.616 Million cell updates/sec

Title: US-10-045-815-4

Perfect score: 3870  
Sequence: 1 MAVRALKLTTLAVVAAS.....PPEPAQESLSGSEKPDDEL 736

Scoring table:

BLOSUM62  
Xgapop 10.0 , Xgapext 0.5  
Ygapop 10.0 , Ygapext 0.5  
Fgapop 6.0 , Fgapext 7.0  
Delop 6.0 , Delext 7.0

Searched: 3190992 seqs, 2439311697 residues

Total number of hits satisfying chosen parameters: 6381984

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%  
Listing first 45 summaries

Command line parameters:

-MODEL=frame+p2n.model p/US10045815/runat 14072004 123009 16861/app query.fasta 1.903  
-DB=/cgn2\_1/USPTO.spool p/US10045815 -SUFFIX=std.rnpb -MINMATCH=0.1  
-DB=Published Applications NA -QWTF=fastap -START=1 -END=1 -MATRIX=blosum62  
-LOOPEL=0 -LOOPEXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=blosum62  
-TRANS=human40.cdi -LIST=45 -DOCALIGN=200 -THR SCORE=pct -THR MAX=100  
-THR\_MIN=0 -ALIGN=15 -MODES=LOCAL -OUTFM=pcp -NOR=ext -HEAPSIZE=500 -MINLEN=0  
-MAXLEN=2000000000 -USER=US10045815 @CGN 1 1 912 @runat 14072004 123009 16861  
-NCPU=6 -ICPU=3 -NO MAP -LARGEQUERY -NEG SCORES=0 -WAIT -DSPBLOC=100  
-LONGLOC -DEV TIMEOUT=120 -WARG TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5  
-FGAPOP=6 -FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : Published Applications NA:

1: /cgn2\_6/ptodata/2/pubpna/US07\_PUBCOMB.seq:  
2: /cgn2\_6/ptodata/2/pubpna/PCT\_NEW\_PUB.seq:  
3: /cgn2\_6/ptodata/2/pubpna/US06\_NEW\_PUB.seq:  
4: /cgn2\_6/ptodata/2/pubpna/US06\_PUBCOMB.seq:  
5: /cgn2\_6/ptodata/2/pubpna/US07\_NEW\_PUB.seq:  
6: /cgn2\_6/ptodata/2/pubpna/PCTUS\_PUBCOMB.seq:  
7: /cgn2\_6/ptodata/2/pubpna/US08\_NEW\_PUB.seq:  
8: /cgn2\_6/ptodata/2/pubpna/US08\_PUBCOMB.seq:  
9: /cgn2\_6/ptodata/2/pubpna/US09A\_PUBCOMB.seq:  
10: /cgn2\_6/ptodata/2/pubpna/US09B\_PUBCOMB.seq:  
11: /cgn2\_6/ptodata/2/pubpna/US09C\_PUBCOMB.seq:  
12: /cgn2\_6/ptodata/2/pubpna/US09\_NEW\_PUB.seq:  
13: /cgn2\_6/ptodata/2/pubpna/US09\_NEW\_PUB.seq2:  
14: /cgn2\_6/ptodata/2/pubpna/US10A\_PUBCOMB.seq:  
15: /cgn2\_6/ptodata/2/pubpna/US10B\_PUBCOMB.seq:  
16: /cgn2\_6/ptodata/2/pubpna/US10C\_PUBCOMB.seq:  
17: /cgn2\_6/ptodata/2/pubpna/US10\_NEW\_PUB.seq:  
18: /cgn2\_6/ptodata/2/pubpna/US10\_NEW\_PUB.seq2:  
19: /cgn2\_6/ptodata/2/pubpna/US60\_PUBCOMB.seq:

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length DB	ID	Description
------------	-------	-------------	-----------	----	-------------

	1	3870	100.0	2600	14	US-10-045-815-4	Sequence 3, Appli
	2	3861	99.8	2242	13	US-10-302-172-228	Sequence 228, App
	3	3846	99.4	2211	13	US-10-257-174-2	Sequence 2, Appli
	4	3846	99.4	2583	13	US-10-312-352-69	Sequence 69, Appl
	5	3801	98.2	2753	9	US-09-728-952-28	Sequence 28, Appl
	6	3801	98.2	2753	9	US-09-728-952-28	Sequence 51, Appl
	7	3780	97.7	2829	14	US-10-045-815-1	Sequence 1, Appli
	8	3645	94.2	2152	13	US-10-302-172-227	Sequence 237, App
	9	3572	92.3	2127	13	US-10-257-174-1	Sequence 1, Appli
	10	3323	85.9	2416	14	US-10-045-815-5	Sequence 5, Appli
	11	3120	80.6	2322	14	US-10-045-815-7	Sequence 7, Appli
	12	2147	55.5	1437	13	US-10-302-172-226	Sequence 226, App
	13	1751	45.2	3396	16	US-10-094-749-921	Sequence 921, App
	14	1715.5	44.3	2665	14	US-10-071-766-86	Sequence 86, Appl
	15	1473	38.1	2305	16	US-10-094-886-23	Sequence 23, Appl
	16	1023.5	26.4	1629	16	US-10-062-674-1793	Sequence 1793, Ap
	17	906.5	23.4	849	16	US-10-012-697-1401	Sequence 1401, Ap
	18	836.5	21.6	1023	9	US-09-764-868-123	Sequence 123, App
	19	836.5	21.6	1023	11	US-09-764-875-508	Sequence 508, App
	20	804	20.8	1849	11	US-09-764-875-508	Sequence 217, App
	21	645	16.7	796	13	US-10-240-425-574	Sequence 574, App
	22	588	15.2	337	13	US-10-085-783A-6419	Sequence 6419, Ap
	23	577	14.9	3871	9	US-10-242-535A-6419	Sequence 1, Appli
	24	501	12.9	2347	9	US-09-962-832-113	Sequence 113, App
	25	501	12.9	2347	16	US-10-439-388-64	Sequence 64, Appl
	26	501	12.9	2443	16	US-10-094-749-724	Sequence 724, App
	27	501	12.9	465	10	US-09-918-995-29327	Sequence 29327, A
	28	492	12.7	465	10	US-10-108-260A-2058	Sequence 2058, Ap
	29	465.5	12.0	1810	16	US-10-108-260A-2058	Sequence 217, App
	30	383.5	9.9	463	13	US-10-276-774-217	Sequence 618, App
	31	381	9.8	340	16	US-10-062-674-618	Sequence 942, App
	32	365.5	9.4	987	12	US-09-876-143-943	Sequence 943, App
	33	365.5	9.4	987	12	US-09-876-143-943	Sequence 943, App
	34	318	8.2	476	9	US-09-783-590-3847	Sequence 3847, Ap
	35	212.5	5.5	2198	16	US-10-108-260A-1485	Sequence 1485, Ap
	36	211.5	5.5	396	13	US-10-085-783A-9785	Sequence 9785, Ap
	37	211.5	5.5	396	16	US-10-242-535A-9785	Sequence 9785, Ap
	38	201	5.2	180	13	US-10-085-783A-2337	Sequence 2337, Ap
	39	201	5.2	180	16	US-10-242-535A-2337	Sequence 2337, Ap
	40	194.5	5.0	301	9	US-09-759-143-271	Sequence 271, App
	41	194.5	5.0	301	9	US-09-780-669-271	Sequence 271, App
	42	194.5	5.0	301	9	US-09-823-827-271	Sequence 271, App
	43	194.5	5.0	301	9	US-09-232-880-271	Sequence 271, App
	44	194.5	5.0	301	9	US-08-895-793-271	Sequence 271, App
	45	194.5	5.0	301	9	US-09-895-814-271	Sequence 271, App

#### ALIGNMENTS

#### RESULT 1

US-10-045-815-3  
; Sequence 3, Application US/10045815  
; Publication No. US20020160498A1  
; GENERAL INFORMATION:  
; APPLICANT: Wadhw, Renu  
; APPLICANT: Sugihara, Takashi  
; APPLICANT: Ohide, Akiko  
; TITLE OF INVENTION: TUMOR SUPPRESSOR GENE  
; FILE REFERENCE: 06501-091001  
; CURRENT APPLICATION NUMBER: US/10/045,815  
; CURRENT FILING DATE: 2001-10-26  
; PRIOR APPLICATION NUMBER: PCT/JP00/02731  
; PRIOR FILING DATE: 2000-04-26  
; PRIOR APPLICATION NUMBER: JP 11/118806  
; PRIOR FILING DATE: 1999-04-26  
; NUMBER OF SEQ ID NOS: 15  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 3  
; LENGTH: 2600  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
; FEATURE:

```
/ NAME/KEY: CDS
/ LOCATION: (52) ... (2259)
US-10-045-815-3
Alignment Scores:
Pred. No.: 0 Length: 2600
Score: 3870.00 Matches: 736
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 100.00% Indels: 0
DB: 14 Gaps: 0

US-10-045-815-4 (1-736) x US-10-045-815-3 (1-2600)

Qy 1 MetAlaValArgAlaLeuLysLeuLeuThrThrLeuLeuAlaValAlaAlaSer 20
Db 52 ATGCGCGGTACGCGGTTGAAGCTGCTGACACACTGCTGGCTGCTGGCGGCTGCTCC 111
Qy 21 GlnAlaGluValGluSerGluAlaGlyTrpGlyMetValThrProAspLeuLeuPheAla 40
Db 112 CAAGCCGAGGTGAGTCCGAGCGAGGATGGGCAATGGTGACGCTGATCTGCTCTTCGCC 171
Qy 41 GluGlyThrAlaAlaIaIaArgGlyAspTrpProGlyValValLeuSerMetGluArg 60
Db 172 GAGGGGACCGAGCTACGCGCGCGGGGAGCTGGCCCGGGTGTCTTGACATGGAAACGG 231
Qy 61 AlaLeuArgSerArgAlaAlaLeuArgAlaLeuArgLeuArgCysArgThrGlnCysAla 80
Db 232 GCGTGTGGCTCCCGGGGAGCGCTCCGGCGCCCTTGGCTGGCGCTCCGACCCAGTGTGCC 291
Qy 81 AlaAspPheProTrpGluLeuAspProAspTrpSerProSerProAlaGlnAlaSerGly 100
Db 292 GCCGACTTCCGTCGGAGCTGGACCCCGAGCTGTGTCGCCCGCGCCGAGGCTCGGGC 351
Qy 101 AlaGlyAlaLeuArgAspLeuSerPhePheGlyGlyLeuLeuArgAlaAlaCysLeu 120
Db 352 GCCGGCGCTTCGCGGACCTGAGCTTCTTCGGGGGCTTCTGGCGTCCGCTGCTCCTG 411
Qy 121 ArgArgCysLeuGlyProProAlaAlaHisSerLeuSerGluGluMetGluLeuGluPhe 140
Db 412 CGCCGCTGCTTGGGGCGCGCGCGCCGCTGCTGCTCAGCGAAGATGGAGCTGGAGTTC 471
Qy 141 ArgLysArgSerProTrpAsnTrpLeuGlnValAlaTrpPheLysLeuAsnLysLeuGlu 160
Db 472 CGCAAGCGGAGCCCTCAACTACTCGAGTCCGCTTCTCAAGATCAACAAGTTGGAG 531
Qy 161 LysAlaValAlaAlaHisThrPhePheValGlyAsnProGluHisMetGluMetGln 180
Db 532 AAAGCTGTGCTGCAGCACACACTTCTTCTGGGGCAATCTGAGCACATGGAAATGCAG 591
Qy 181 GlnAsnLeuAspTrpTrpGlnThrMetSerGlyValLysGluAlaAspPheLysAspLeu 200
Db 592 CAGAACTAGACTATTACCAAAACCATCTCTGGAGTGAAGCGCGGACTTCAAGGATCTT 651
Qy 201 GluThrGlnProHisMetGlnGluPheArgLeuGlyValArgLeuTrpSerGluGluGln 220
Db 652 GAGACTCAACCCCATATGCAAGAAATTCGACTGGGAGTGGCTACTCAGAGGAACAG 711
Qy 221 ProGlnGluAlaValProHisLeuGluAlaLeuGlnGluTrpPheValAlaTrpGlu 240
Db 712 CCACAGGAGCTGTGCCCCACCTAGAGCGCGGCTGCAAGAAATACITTTGGGCTATGAG 771
Qy 241 GluCysArgAlaLeuCysGluGlyProTrpAspTrpAspGlyTrpAsnTrpLeuGluTrp 260
Db 772 GAGTGGCGTGGCTCTCGAAGGCGCCCTATGACTACGATGGCTACCACTACCTTGGATAC 831
Qy 261 AsnAlaAspLeuPheGlnAlaIleThrAspHisTrpIleGlnValLeuAsnCysLysGln 280
Db 832 AACCTGACCTCTTCCAGGGCAATCACAGATCATATACATCCAGTCTCCTCACTGTAGCAG 891
Qy 281 AsnCysValThrGluLeuAlaSerHisProSerArgGluLysProPheGluAspPheLeu 300
Db 892 AACTGTGTACGAGCTGTGCTCCACCCCAAGTCGAGAGAAGCCCTTTGAAGACTTCTCCTC 951
```

```
Qy 301 ProSerHisTrpAsnTrpLeuGlnPheAlaTrpTrpAsnIleGlyAsnTrpThrGlnAla 320
Db 952 CCATCGCATTAATTAATCTGCGTTTGCCTACTATAACATTTGGAAATTAACAAAGCT 1011
Qy 321 GlyGluCysAlaLysThrTrpLeuLeuPhePheProAsnAspGluValMetAsnGlnAsn 340
Db 1012 GGTGAATGTGCAAGACCTATCTTCTTCTTCCCAATGACGAGGTGATGAACCAAAAT 1071
Qy 341 LeuAlaTrpTrpAlaAlaMetLeuGlyGluGluHisThrArgSerIleGlyProArgGlu 360
Db 1072 TTGGCTTATTATGACAGTATGCTTGGAGAGAACACACACAGATCCATCGGCCCGCTGAG 1131
Qy 361 SerAlaLysGluTrpArgGlnArgSerLeuLeuLysGluLeuLeuPhePheAlaTrp 380
Db 1132 AGTGCACAGGAGTACCGACAGCAAGCTACTGAAAAGAACTGCTTTCTTCCGCTTAT 1191
Qy 381 AspValPheGlyIleProPheValAspProAspSerTrpThrProGluGluValIlePro 400
Db 1192 GATGTTTGGTAATTCCTTTGTGATCCGATTCATGGACTCCAGAGAAGTATGATCCC 1251
Qy 401 LysArgLeuGlnGluLysGlnLysSerGluArgGluThrAlaValAlaArgIleSerGlnGlu 420
Db 1252 AAGAGATTGCAAGAGAAACAGAAAGTCAGAACGGGAAACAGCGCTACGCATCTCCAGGAG 1311
Qy 421 IleGlyAsnLeuMetLysGluIleGluThrLeuValGluGluLysThrLysGluSerLeu 440
Db 1312 ATTGGGAACCTTATGAGAAATCGAGACCTTGTGGAGAGAAAGACCAAGGAGTCACTG 1371
Qy 441 AspValSerArgLeuThrArgGluGlyGlyProLeuLeuTrpGluGlyIleSerLeuThr 460
Db 1372 GATGTGACGAGACTGACCCGGGAGGTGGCCCTGCTGTATGAAGCATCAGTCTCAC 1431
Qy 461 MetAsnSerLysLeuLeuAsnGlyTrpGlnArgValValMetAspGlyValIleSerAsp 480
Db 1432 ATGAATCCAAACTCCTGAATGGTTACGAGCGGTGGTATGGACGGCGTNAATCTCTGAC 1491
Qy 481 HisGluCysGlnGluLeuGlnArgLeuThrAsnValAlaAlaThrSerGlyAspGlyTrp 500
Db 1492 CACGAGTGTGAGGAGTGCAGAGACTGACCAATGTGGCAGCAACTCAGGAGATGGCTAC 1551
Qy 501 ArgGlyGlnThrSerProHisThrProAsnGluLysPheTrpGlyValThrValPheLys 520
Db 1552 CGGGGTGAGACTCCCCACATCTCCCAATGAAAGTTCTATGGTGTCTACTGTCTTCAAA 1611
Qy 521 AlaLeuLysLeuGlyGlnGluLysValProLeuGlnSerAlaHisLeuTrpTrpAsn 540
Db 1612 GCCTCAAGCTGGGGCAGAGGCAAGTTCTCTGCAGAGTGGCCACCTGTACTACAAC 1671
Qy 541 ValThrGluLysValArgArgIleMetGluSerTrpPheArgLeuAspThrProLeuTrp 560
Db 1672 GTGACGAGAAAGTGGCGGCTCATCGAGTCTTACTTCCGCTGGATACGCCCTCTAC 1731
Qy 561 PheSerTrpSerHisLeuValCysArgThrAlaIleGluGluValGlnAlaGluArgLys 580
Db 1732 TTTTCTTCTTCTATCTGTGTGTGCGCACTGCCATCGAAGAGGTCCAGGAGAGAGAG 1791
Qy 581 AspAspSerHisProValHisValAspAsnCysIleLeuAsnAlaGluThrLeuValCys 600
Db 1792 GATGATAGTATCCAGTCCAGTCCAGCAACTGCATCTCTGAATGCGGAGACCTCGTGTGT 1851
Qy 601 ValLysGluProProAlaTrpThrPheArgAspTrpSerAlaIleLeuTrpLeuAsnGly 620
Db 1852 GTCAAGAGCCCCCAGCCCTACACCTTCCGCGACTACAGCGCATCTTACCTTAATGGG 1911
Qy 621 AspPheAspGlyGlyAsnPheTrpThrGluLeuAspAlaLysThrValThrAlaGlu 640
Db 1912 GACTTCGATGCGGAAACTTTTATTTTCACTGAATGGATGCAAGACCGTGCAGCGCAGAG 1971
Qy 641 ValGlnProGlnCysGlyArgAlaValGlyPheSerSerGlyThrGluAsnProHisGly 660
Db 1972 GTGCAGCCTCAGTGTGGAAGAGCGGTGGATTCTTCTCAGGCATGAAACCCACATGGA 2031
```

QY 661 ValIysAlaValThrArgGlyGlnArgCysAlaIleAlaLeuTrpPheThrLeuAspPro 680  
DB 2032 GTGAAGGCTGTACACAGGGGAGCGCTGTGCCATCGCCCTGTGGTTACCCCTGGACCT 2091  
QY 681 ArgHisSerGluArgAspArgValGlnAlaAspAspLeuValIysMetLeuPheSerPro 700  
DB 2092 CGACACAGCGAGCGGACAGGGTGCAGGAGATGACCTGTGGAAGATGCTTTTCAGCCCA 2151  
QY 701 GluGluMetAspLeuSerGlnGluGlnProLeuAspAlaGlnGlnGlyProProGluPro 720  
DB 2152 GAAGAGATGACCTCTCTCCAGGAGCAGCCCTGGATGCCACGAGGGCCCCCGAACCT 2211  
QY 721 AlaGlnGluSerLeuSerGlySerGluSerIysProIysAspGluLeu 736  
DB 2212 GCACAGAGTCTCTCTCAGGAGTGAATCGAAGCCCAAGGATGAGCTA 2259  
RESULT 2  
US-10-302-172-228  
; Sequence 228, Application US/10302172  
; Publication No. US20040053250A1  
; GENERAL INFORMATION:  
; APPLICANT: Tang, Y. Tom  
; APPLICANT: Xue, Aidong J.  
; APPLICANT: Drmanac, Radoje T.  
; TITLE OF INVENTION: No. US20040053250A1el Arginine-rich Protein-like Nucleic Acids an  
; FILE REFERENCE: 803\_1CNCp  
; CURRENT APPLICATION NUMBER: US/10/302,172  
; CURRENT FILING DATE: 2002-11-21  
; PRIOR APPLICATION NUMBER: US 10/225,251  
; PRIOR FILING DATE: 2002-08-20  
; PRIOR APPLICATION NUMBER: PCT US02/05095  
; PRIOR FILING DATE: 2002-03-05  
; PRIOR APPLICATION NUMBER: US 09/799,451  
; PRIOR FILING DATE: 2001-03-05  
; NUMBER OF SEQ ID NOS: 950  
; SOFTWARE: pt\_FL\_genes Version 2.0  
; SEQ ID NO 228  
; LENGTH: 2242  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
; FEATURE:  
; NAME/KEY: CDS  
; LOCATION: (32)..(2239)  
US-10-302-172-228  
Alignment Scores:  
Pred. No.: 0 Length: 2242  
Score: 3861.00 Matches: 734  
Percent Similarity: 99.86% Conservative: 1  
Best Local Similarity: 99.73% Mismatches: 1  
Query Match: 99.77% Indels: 0  
DB: 13 Gaps: 0  
US-10-045-815-4 (1-736) x US-10-302-172-228 (1-2242)  
QY 1 MetAlaValArgAlaLeuIysLeuLeuThrThrLeuLeuAlaValAlaAlaLaser 20  
DB 32 ATGGCGGTACGCGGTGAGCTGTGACCACTGCTGGTGTCTGGTGGCCCTGGCTCC 91  
QY 21 GlnAlaGluValGluSerGluAlaGlyTrpGlyMetValThrProAspLeuLeuPheAla 40  
DB 92 CAAGCCGAGTCCGAGTCCGAGGAGGATGGGCAATGGTCACTGATCTGCTCTTCGCC 151  
QY 41 GluGlyThrAlaAlaValArgGlyAspTrpProGlyValValLeuSerMetGluArg 60  
DB 152 GAGGAGCCCGACCTACTTCGCGGGGAGTGGCCGGGGTGGTCTCGAGCATGGAAACGG 211  
QY 61 AlaLeuArgSerArgAlaAlaLeuArgAlaLeuArgLeuArgCysArgThrGlnCysAla 80  
DB 212 GCGCTCGCTCCGCGGAGCCCTTCGCGCCCTTCGCTGGCTGGCGCACCAGTGTGCC 271  
QY 81 AlaAspPheProTrpGluLeuAspProAspTrpSerProAlaGlnAlaLaserGly. 100

DB 272 GCGGACTTCCGTGGAGCTGGACCCGACTGGTCCCGACCGCCGCGCCAGGCTTCGGGC 331  
QY 101 AlaGlyAlaLeuArgAspLeuSerPhePheGlyGlyLeuLeuArgArgAlaAlaCysLeu 120  
DB 332 GCGCGCGCGCTGCGGAGCTTCTGCGGGGCGCTTCTGCGTGGCGCTGCGCTGCGCTG 391  
QY 121 ArgArgCysLeuGlyProProAlaAlaHisSerLeuSerGluGluMetGluLeuGluPhe 140  
DB 392 CGCGCTGCTCGGCG 451  
QY 141 ArgIysArgSerProTyrAsnTyrLeuGlnValAlaIleTyrPheIysIleAsnIysLeuGlu 160  
DB 452 CGCAAGCGGAGCGCCCTACAACTACCTGCGAGGTGCGCTACTTCAAGATCAACAAGTTGAG 511  
QY 161 LysAlaValAlaAlaAlaHisThrPheValGlyAsnProGluHisMetGluMetGln 180  
DB 512 AAAGCTGTGTGTCGAGCACACACTTCTTCTGGGCAATCTCTGACCATGATGAAATGCG 571  
QY 181 GlnAsnLeuAspTyrTyrGlnThrMetSerGlyValIysGluAlaAspPheIysAspLeu 200  
DB 572 CAGAACCTAGACTATTACCAAAACATGTCTGGAGTGAAGGAGGCGGACTTCAAGGATCTT 631  
QY 201 GluThrGlnProHisMetGlnGluPheArgGlyValArgLeuTyrSerGluGluGln 220  
DB 632 GAGACTCAACCCCATATCAAGAAATTTGACTGGGAGTGGGACTCTACTCAGAGGAAACAG 691  
QY 221 ProGlnGluAlaValProHisGluAlaAlaLeuGlnGluTyrPheValAlaIleTyrGlu 240  
DB 692 CCACAGGAGCTGTGCCCCACCTAGAGCGCGGCTGCGAGAAATACTTGTGGCTATGAG 751  
QY 241 GluCysArgAlaLeuCysGluGlyProTyrAspTyrAspGlyTyrAsnTyrLeuGluTyr 260  
DB 752 GAGTCCGCTGCGCTCTGCGAAGGCGCTTATGACTACGATGCGTACAACTACCTTGGTAC 811  
QY 261 AsnAlaAspLeuPheGlnAlaIleThrAspHisTyrIleGlnValLeuAsnCysIysGln 280  
DB 812 AACGCTGACCTCTTCCAGGCCATCACAGATCATTCATCCAGGCTCTCATCTGTAAAGCAG 871  
QY 281 AsnCysValThrGluLeuAlaSerHisProSerArgGluIysProPheGluAspPheLeu 300  
DB 872 AACTGTGTACGGAGCTTGTCTCCACCAAGTCAGAGAAAGCCCTTTGAAGACTTCCCTC 931  
QY 301 ProSerHisTyrAsnTyrLeuGlnPheAlaIleTyrAsnIleGlyAsnTyrThrGlnAla 320  
DB 932 CCATCGCATTAATATTATCTGAGTTTGCCTACTATACATTTGGGAATTTATACAAAGCT 991  
QY 321 GlyGluCysAlaIysThrTyrLeuLeuPhePheProAsnAspGluValMetAsnGlnAsn 340  
DB 992 GGTGAATGTGCAAGACCTATCTTCTTCTTCCCAATGACGAGGTGATGAACCAAAAT 1051  
QY 341 LeuAlaIleTyrAlaAlaMetLeuGlyGluGluHisThrArgSerIleGlyProArgGlu 360  
DB 1052 TTGGCTATTATGAGCTATGCTTTGGAGAAACACACCAAGATCCATCGGCCCCCGTGAG 1111  
QY 361 SerAlaIysGluTyrArgGlnArgSerLeuLeuGluIysGluLeuLeuPheAlaIleTyr 380  
DB 1112 AGTGCAAGGAGTACCGACGAGGAGCTACTGGAAAAGAACTGCTTTTCTTGGCTTAT 1171  
QY 381 AspValPheGlyIleProPheValAspProAspSerTrpThrProGluGluValIlePro 400  
DB 1172 CATGTTTTTGGAAATTCCTTTTGTGGATCCGATTCATGGACTCCAGAGAAAGTGAATCCC 1231  
QY 401 LysArgLeuGlnGluIysGlnIysSerGluArgGluThrAlaValArgIleSerGlnGlu 420  
DB 1232 AAGAGATTGCAAGAGAAACAGAGTCAGAACGGGAAACAGCCGACCGATCTCCAGGAG 1291  
QY 421 IleGlyAsnLeuMetLysGluIleGluThrLeuValGluGluIysThrLysGluSerLeu 440  
DB 1292 ATTGGAACTTATGAAGAAATCGAGACCTTGTGGAGAGAGAACCAAGGAGTCACTG 1351  
QY 441 AspValSerArgLeuThrArgGluGlyIleProLeuLeuTyrGluIleSerLeuThr 460



```
Db 1352 GATGTGACGAGACTACCGGGAGGTGGCCCTGCTGTATGAAGGCATCAGTCTCACC 1411
Qy 461 MetAsnSerLysLeuLeuAsnGlyTyrGlnArgValValMetAspGlyValIleSerAsp 480
Db 1412 ATGACTCCAACTCCTGANTGTTACCAAGGGTGGTGTGATGACGGCGGTAACTCTGAC 1471
Qy 481 HisGluCysGlnGluLeuGlnArgLeuThrAsnValAlaAlaThrSerGlyAspGlyTyr 500
Db 1472 CACGAGTGTGAGGAGTGCAGAGACTGACCAATGTGGCAGCAACCTCAGGAGATGGCTAC 1531
Qy 501 ArgGlyGlnThrSerProHisThrProAsnGluLysPheTyrGlyValThrValPheLys 520
Db 1532 CGGGGTGAGACCTCCCACTACTCCCAATGAAGAATTCATGTGTCTACTGTCTTCAA 1591
Qy 521 AlaLeuLysLeuGlyGlnGluGlyLysValProLeuGlnSerAlaHisLeuTyrTyrAsn 540
Db 1592 GCCCTCAAGCTGGGGCAAGGCAAGTTCCTCTGCAGAGTGCACCACTGTACTACAAC 1651
Qy 541 ValThrGluLysValArgArgIleMetGluSerTyrPheArgLeuAspThrProLeuTyr 560
Db 1652 GTGACGGAGAAAGTGGCGGCATCATGGAGTCTCTTCCGCTGGATACGCCCTCTAC 1711
Qy 561 PheSerTyrSerHisLeuValCysArgThrAlaIleGluValGlnAlaGluArgLys 580
Db 1712 TTTTCTCTACTCTCATCTGGTGTGGCGCACTGCCATCGAAGAGTCCAGGAGAGGAAG 1771
Qy 581 AspAspSerHisProValHisValAspAsnCysIleLeuAsnAlaGluThrLeuValCys 600
Db 1772 GATGATAGTCATCCAGTCCACGTGGACAACTGATCCTGTAATCCGAGACCTCTGTGT 1831
Qy 601 ValLysGluProProAlaTyrThrPheArgAspTyrSerAlaIleLeuTyrLeuAsnGly 620
Db 1832 GTCAAGAGAGCCCGAGCTACACTTCCGCGACTACAGCGCCATCTTTACCTAAATGGG 1891
Qy 621 AspPheAspGlyGlyAsnPheTyrPheThrGluLeuAspAlaLysThrValThrAlaGlu 640
Db 1892 GACTTCGATGGCGAAACTTTATTTTCACTGAATGGATGCCAAGACCGGTGACGGCAGAG 1951
Qy 641 ValGlnProGlnCysGlyArgAlaValGlyPheSerSerGlyThrGluAsnProHisGly 660
Db 1952 GTGCAGCTCAGTGTGAAGAGCGGTGGGATTTCTTTTCAGGCACCTGAAAACCCCAATGA 2011
Qy 661 ValLysAlaValThrArgGlyGlnArgCysAlaIleAlaLeuThrPheThrLeuAspPro 680
Db 2012 GTGAAGGCTGTTCACAGGGGGCAGCGCTGTGCCATCGCCCTGTGGTTCACCTGGACCT 2071
Qy 681 ArgHisSerGluArgAspArgValGlnAlaAspAspLeuValLysMetLeuPheSerPro 700
Db 2072 CGACACAGCAGCGGGACAGGGTGCAGGAGATGACCTGTGTGAAGATGCTCTTCAGCCCA 2131
Qy 701 GluGluMetAspLeuSerGlnGluGlnProLeuAspAlaGlnGlnGlyProProGluPro 720
Db 2132 GAAGAGATGACCTCTCCAGGAGCAGCCCTCGATGTCGCCAGCGGGCCCCCCCCGAACCT 2191
Qy 721 AlaGlnGluSerLeuSerGlySerGluSerLysPheProLysAspGluLeu 736
Db 2192 GCACAAGAGTCTCTCTCAGGAGTGAATCGAAGCCCAAGGATGAGCTA 2239
```

## RESULT 3

```
US-10-257-174-2
; Sequence 2, Application US/10257174
; Publication No. US20040034194A1
; GENERAL INFORMATION:
; APPLICANT: Agarwal, Pankaj
; APPLICANT: Murdock, Paul R.
; APPLICANT: Rizvi, Safia K.
; APPLICANT: Smith, Randall F.
; APPLICANT: Xiang, Zhaoying
; TITLE OF INVENTION: NOVEL COMPOUNDS
; FILE REFERENCE: GP50022
; CURRENT APPLICATION NUMBER: US/10/257,174
; CURRENT FILING DATE: 2002-10-10
; PRIOR APPLICATION NUMBER: PCT/US01/11797
```

```
; PRIOR FILING DATE: 2001-04-11
; PRIOR APPLICATION NUMBER: 60/196,603
; PRIOR FILING DATE: 2000-04-13
; PRIOR APPLICATION NUMBER: 60/199,417
; PRIOR FILING DATE: 2000-04-24
; NUMBER OF SEQ ID NOS: 48
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 2
; LENGTH: 2211
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-257-174-2
```

```
Alignment Scores:
Pred. No.: 0 Length: 2211
Score: 3846.00 Matches: 733
Percent Similarity: 99.59% Conservative: 0
Best Local Similarity: 99.59% Mismatches: 3
Query Match: 99.38% Indels: 0
DB: 13 Gaps: 0
```

US-10-045-815-4 (1-736) x US-10-257-174-2 (1-2211)

```
Qy 1 MetAlaValArgAlaLeuLysLeuLeuThrThrLeuLeuAlaValAlaAlaSer 20
Db 1 ATGGCGGTACGCGCTTGAAGCTGCTGACCACTGCTGGCTGTCTGGCGCTGCTCC 60
Qy 21 GlnAlaGluValGluSerGluAlaGlyTyrGlyMetValThrProAspLeuLeuPheAla 40
Db 61 CAAGCCGAGGTGCGATCCGAGCGAGATGGGGCATGGTGTGAGCGCTGATCTCTTCGCC 120
Qy 41 GluGlyThrAlaAlaTyrAlaArgGlyAspTyrProGlyValValLeuSerMetGluArg 60
Db 121 GAGGGGACCGCAGCTACGCGCGGGGACTGGCCCGGGGTGGTCTCTGAGCATGGAACGG 180
Qy 61 AlaLeuArgSerArgAlaAlaLeuArgAlaLeuArgLeuArgCysArgThrGlnCysAla 80
Db 181 GCGCTGCGCTCCCGGCGAGCCCTCCGCGCCCTTCGCTCGCGCTCCGCGCCAGCATGTGCC 240
Qy 81 AlaAspPheProTyrGluLeuAspProAspTyrSerProSerProAlaGlnAlaSerGly 100
Db 241 GCCGACTTCCCGTGGAGCTGGACCCCGACTGTGTCGCCAGCCCGCCAGCCCTCGGGC 300
Qy 101 AlaGlyAlaLeuArgAspLeuSerPhePheGlyGlyLeuLeuArgArgAlaAlaCysLeu 120
Db 301 GCGCGCGCCCTGCGCGACCTGAGCTTCTTCGGGGGCGCTTCTGCGCTCGCGCTGCGCTG 360
Qy 121 ArgArgCysLeuGlyProProAlaAlaHisSerLeuSerGluGluMetGluLeuGluPhe 140
Db 361 GCGCGCTCCCTCGGGCGCCCGCGCCCGCCACTGCTCGCGAGAGATGGAGCTGGAGTTC 420
Qy 141 ArgLysArgSerProTyrAsnTyrLeuGlnValAlaTyrPheLysIleAsnLysLeuGlu 160
Db 421 CGAAGCGAGGAGCCCTACCACTTACCTGAGGTGCGCTTCTTCAAGATCAACAAGATTGGAG 480
Qy 161 LysAlaValAlaAlaAlaHisThrPhePheValGlyAsnProGluHisMetGluMetGln 180
Db 481 AAAGCTGTGTGCAGCAGCACACCTTCTTCGTGGGCAATCCTGAGCAGCATGGAAATGCAG 540
Qy 181 GlnAsnLeuAspTyrTyrGlnThrMetSerGlyValLysGluAlaAspPheLysAspLeu 200
Db 541 CAGAACCTAGACTATTACCAACCACTGCTCTGAGTGAAGGAGGCCGACTTCAAGATCTT 600
Qy 201 GluThrGlnProHisMetGlnGluPheArgLeuGlyValArgLeuTyrSerGluGluGln 220
Db 601 GAGACTCAACCCCATATGCAAGAAATTTTCGACTGGGAGTGGGACTCTACTCAGAGGAACAG 660
Qy 221 ProGlnGluAlaValProHisLeuGluAlaAlaLeuGlnGluTyrPheValAlaTyrGlu 240
Db 661 CCACAGGAGGTGTGCCCCACCTAGAGGGCGGTGCAAGATATCTTTGTGGCTATGAG 720
Qy 241 GluCysArgAlaLeuCysGluGlyProTyrAspTyrAspGlyTyrAsnTyrLeuGlyTyr 260
```



721 Db GAGTGGCGTCCCTCTGCGAAGGGCCCTATGACTACGATGGCTACAACTACCTTGAGTAC 780  
261 Qy AsnAlaAspLeuPheGlnAlaIleThrAspHisTyrIleGlnValLeuAsnCysLysGln 280  
781 Db AACGCTGACCTCTTCCAGGCCATCAGATCATTAATCAGGTCCTCAACTGTGAAGCAG 840  
281 Qy AsnCysValThrGluLeuAlaSerHisProSerArgLysProPheGluAspPheLeu 300  
841 Db AACGTGTCTACGAGGTGTCTTCCCAAGTGTGAGAGAGCCCTTTGAAGACTTCCTC 900  
301 Qy ProSerHisTyrAsnTyrLeuGlnPheAlaTyrTyrAsnIleGlyAsnTyrThrGlnAla 320  
901 Db CCATCGCATTAATTAATCTGCGAGTTTGCCTACTATAACATTGGGAATTATATACAGGCT 960  
321 Qy GlyGluCysAlaLysThrTyrLeuLeuPhePheProAsnAspGluValMetAsnGlnAsn 340  
961 Db GTTGAATGTGCCAAGACTATCTTCTCTTCCCAATGACGAGGTGATGAACCAAAAT 1020  
341 Qy LeuAlaTyrTyrAlaAlaMetLeuGlyGluGluHisThrArgSerIleGlyProArgGlu 360  
1021 Db TTGGCTTATTATGAGCTATGCTGTGAGAGAAACACACACAGATCCATCGGCCCCCGTGAG 1080  
361 Qy SerAlaLysGluTyrArgGlnArgSerLeuLeuGluLysGluLeuLeuPheAlaTyr 380  
1081 Db AGTGGCAAGAGTACCGACAGCGAGCTACTGGAAAAGAACTGCTTTTTCGCTTAT 1140  
381 Qy AspValPheGlyIleProPheValAspProAspSerTyrThrProGluGluValIlePro 400  
1141 Db GATGTTTTTGGAAATTCCTTGTGATTCGGATTCATGACTCCAGAGAAAGTATTCCTC 1200  
401 Qy LysArgLeuGlnGluLysGlnLysSerGluArgGluThrAlaValArgIleSerGlnGlu 420  
1201 Db AAGAGATGCAAGAGAAACAGAGTCAAGACGGGAACACCCCTGATCGCATCTCCAGGAG 1260  
421 Qy IleGlyAsnLeuMetLysGluIleGluThrLeuValGluGluLysThrLysGluSerLeu 440  
1261 Db ATTGGAAACCTTATGAAGAAATCGAGACCCCTTGTGCAAGAGAAGACCAAGAGTCACTG 1320  
441 Qy AspValSerArgLeuThrArgGluGlyGlyProLeuLeuTyrGluGlyIleSerLeuThr 460  
1321 Db GATGTGACGAGATGACCCCGGAAGGTGGCCCTCTGCTGATGAAGGCATCAGTCTCAC 1380  
461 Qy MetAsnSerLysLeuLeuAsnGlyTyrGlnArgValValMetAspGlyValIleSerAsp 480  
1381 Db ATGACTCCAACTCTCTGATGGTCCACGCGGTGGTATGACGCGGTATCTCTGAC 1440  
481 Qy HisGluCysGlnGluLeuGlnArgLeuThrAsnValAlaAlaThrSerGlyAspGlyTyr 500  
1441 Db CACGAGTGTGAGGAGTGTGACAGACTGACCAATGTGCGACAAACCTCAGGAGATGGCTAC 1500  
501 Qy ArgGlyGlnThrSerProHisThrProAsnGluLysPheTyrGlyValThrValPheLys 520  
1501 Db CGGGGTGAGACCTCCCAATATCTCCCAATGAAGTTCTATGGTGTCACTGTCTTCAA 1560  
521 Qy AlaLeuLysLeuGlyGlnGluGlyLysValProLeuGlnSerAlaHisLeuTyrTyrAsn 540  
1561 Db GCCCTCAAGTGGGGCAAGAGGCAAAAGTTCTCTGAGAGTGCCCACTGTACTACTAAC 1620  
541 Qy ValThrGluLysValArgAlaIleMetGluSerTyrPheArgLeuAspThrProLeuTyr 560  
1621 Db GTGACGAGAGAGGTGGCGGCATCATGAGTCTTCTTCCGCTGGATCGCCCTCTAC 1680  
561 Qy PheSerTyrSerHisLeuValCysArgThrAlaIleGluGluValGlnAlaGluArgLys 580  
1681 Db TTTTCTTACTCTCATCTGTTGTGCGCACTGCCATCGAAGAGTCCAGGCGAGAGGAAG 1740  
581 Qy AspAspSerHisProValHisValAspAsnCysAlaIleLeuAsnAlaGluThrLeuValCys 600  
1741 Db GATGATAGTATCATCAGTCCACGTGGACATCGATCTGATCGGAGACCCCTCGTGTGT 1800  
601 Qy ValLysGluProProAlaTyrThrPheArgAspTyrSerAlaIleLeuTyrLeuAsnGly 620  
1801 Db GTCAAAGAGCCCCAGGCTACACCTTCCGCGACTACAGCGCCATCTTTACTTAATGGG 1860

## RESULT 4

US-10-312-352-69  
; Sequence 69, Application US/10312352  
; Publication No. US20040053824A1  
; GENERAL INFORMATION:  
; APPLICANT: INCYTE GENOMICS, INC.; TANG, Y. Tom  
; APPLICANT: YUE, Henry; AZIMZAI, Yalda  
; APPLICANT: HE, Ann; BATRA, Sajeev  
; APPLICANT: LO, Terence P.; NGUYEN, Danniell B.  
; APPLICANT: BURRILL, John D.; MARCUS, Gregory A.  
; APPLICANT: ZINGLER, Kurt A.; GANDHI, Ameena R.  
; APPLICANT: LAL, Preeti G.; KEARNEY, Liam  
; APPLICANT: BURFORD, Neil; YAO, Monique G.  
; APPLICANT: CHAWLA, Narinder K.; ELLIOT, Vicki S.  
; APPLICANT: ARVIZU, Chandra S.; KHAN, Farrah A. J.A.  
; APPLICANT: BOUGHN, Mariah R.; HAFALIA, April, J.A.  
; APPLICANT: POLICK, Jennifer L.; AU-YOUNG, Janice K.  
; APPLICANT: LU, Yan; BOROWSKY, Mark L.  
; APPLICANT: YANG, Junming; GURURAJAN, Rajagopal  
; APPLICANT: WARREN, Bridget A.; GIETZEN, Kimberly J.  
; APPLICANT: XU, Yuming; KALLICK, Deborah A. Kavitha  
; APPLICANT: LEE, Ernestine A.; THANGAVELU, Sally  
; TITLE OF INVENTION: EXTRACELLULAR MATRIX AND CELL ADHESION MOLECULES  
; FILE REFERENCE: PF-0794 USN  
; CURRENT APPLICATION NUMBER: US/10/312,352  
; CURRENT FILING DATE: 2002-12-18  
; PRIOR APPLICATION NUMBER: PCT/US01/21067  
; PRIOR FILING DATE: 2001-06-29  
; PRIOR APPLICATION NUMBER: US 60/215,454  
; PRIOR FILING DATE: 2000-06-30  
; PRIOR APPLICATION NUMBER: US 60/219,462  
; PRIOR FILING DATE: 2000-07-18  
; PRIOR APPLICATION NUMBER: US 60/240,111  
; PRIOR FILING DATE: 2000-10-12  
; PRIOR APPLICATION NUMBER: US 60/240,106  
; PRIOR FILING DATE: 2000-10-12  
; PRIOR APPLICATION NUMBER: US 60/244,021  
; PRIOR FILING DATE: 2000-10-27  
; PRIOR APPLICATION NUMBER: US 60/248,887  
; PRIOR FILING DATE: 2000-11-14  
; PRIOR APPLICATION NUMBER: US 60/249,570  
; PRIOR FILING DATE: 2000-11-16  
; NUMBER OF SEQ ID NOS: 72  
; SOFTWARE: PERL Program  
; SEQ ID NO 69  
; LENGTH: 2583



```
Db 1962 GTGAGCCTCAGTGTGAAGAGCGTGGATTCTCTTCAGGCACTGAAGAACCCACATGA 2021
Qy 661 VallyAlaValThrArgGlyGlnArgCysAlaIleAlaLeuTrpPheThrLeuAspPro 680
Db 2022 GTGAAGGCTGTCAACAGGGGACAGCGCTGTGCCATCCCTGTGGTTCACCTCGACCCCT 2081
Qy 681 ArgHisSerGluArgAspArgValGlnAlaAspLeuValLysMetLeuPheSerPro 700
Db 2082 CGACACAGCGAGCGGACAGCGGTGCGAGCGATGACTGTGTGAGATGCTCTTCAGCCCCA 2141
Qy 701 GluGluMetAspLeuSerGlnGluGlnProLeuAspAlaGlnGlnGlyProProGluPro 720
Db 2142 GAAGAGATGACCTCTCCAGGAGCAGCGCTGATGCCAGCAGGCGCCCCCGCAACCT 2201
Qy 721 AlaGlnGluSerLeuSerGlySerGluSerLysProLysAspGluLeu 736
Db 2202 GCACAAAGAGTCTCTCAGGCACTGAATCGAAGCCCAAGATGAGCTA 2249

RESULT 5
US-09-728-952-28
; Sequence 28, Application US/09728952
; Patent No. US20020111302A1
; GENERAL INFORMATION:
; APPLICANT: Tang, Y. Tom
; APPLICANT: Zhou, Ping
; APPLICANT: Goodrich, Ryle
; APPLICANT: Liu, Chenghua
; APPLICANT: Asundi, Vinod
; APPLICANT: Wang, Jian-Rui
; APPLICANT: Wang, Dunrui
; APPLICANT: Yamazaki, Vicki
; APPLICANT: Ujwal, Manusha L.
; APPLICANT: Dmanac, Radoje T.
; TITLE OF INVENTION: No. US20020111302A1el Nucleic Acids and
; FILE OF INVENTION: Polypeptides
; FILE REFERENCE: 799
; CURRENT APPLICATION NUMBER: US/09/728,952
; CURRENT FILING DATE: 2000-11-30
; NUMBER OF SEQ ID NOS: 101
; SOFTWARE: pt_FL_genes Version 2.0
; SEQ ID NO 28
; LENGTH: 2753
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1)..(2418)
US-09-728-952-28

Alignment Scores:
Pred. No.: 0 Length: 2753
Score: 3801.00 Matches: 733
Percent Similarity: 90.94% Conservative: 0
Best Local Similarity: 90.94% Mismatches: 3
Query Match: 98.22% Indels: 70
DB: 9 Gaps: 1

US-10-045-815-4 (1-736) x US-09-728-952-28 (1-2753)
Qy 1 MetAlaValArgAlaLeuLysLeuLeuThrThrLeuLeuAlaValAlaAlaSer 20
Db 1 ATGGCGGTACGCGGTGTGAAGCTGCTGACACACTGCTGGCTGTCTGGCGCTGCTCC 60
Qy 21 GluAlaGluValGluSerGluAlaGlyTrpGlyMetValThrProAspLeuLeuPheAla 40
Db 61 CAAGCCGAGGTCGAGTCCGAGGAGGATGGGCGATGGCGCTGATCTGCTCTTCGCC 120
Qy 41 GluGlyThrAlaAlaValArgGlyAspTrpProGlyValValLeuSerMetGluArg 60
Db 121 GAGGAGACCCAGCTACGCGCGCGGAGCTGCGCGGGGTGTCTCTGAGCATGGHACGG 180
Qy 61 AlaLeuArgSerArgAlaAlaLeuArgAlaLeuArgLeuArgCysArgThrGlnCysAla 80
```

```
181 GCGCTGCGCTCCGGGAGCCCTCCGCGCCCTTCGCTGCGCTCCGCGACCCACGATGTGCC 240
81 AlaAspPheProTrpGluLeuAspProAspTrpSerProSerProAlaGlnAlaSerGly 100
241 GCGGACTTCCGCTGGGAGCTGACCCCGACTGGTCCCCAGCCGCGCCAGCCCTCGGGC 300
101 AlaGlyAlaLeuArgAspLeuSerPhePheGlyGlyLeuLeuArgArgAlaAlaCysLeu 120
301 GCGCGCGCTCCGCGGACCTGAGCTTCTTCGGGGCCCTTCGCTGCGGCTCGCTTCCTG 360
121 ArgArgCysLeuGlyProProAlaAlaHisSerLeuSerGluGluMetGluLeuGluPhe 140
361 GCGCGCTGCTCCGCGCGCCGCGCCCTCGCTCAGCGAAGAGATGGAGCTGGAGTTTC 420
141 ArgLysArgSerProTrpAsnTyrLeuGlnValAlaAlaTyrPhe----- 154
421 CGCAAGCGGAGCCCTCACTCACTGCTGAGTGCCTACTTCAAGGTGCAGACCTGCCTG 480
154 ----- 154
481 GAACCGAGCGCGCGGGTCTTCTGGGAGAGAGAGTGTTCGAGGGGACCTGAGGAGCTTG 540
154 ----- 154
541 GGGGATCGGGGAAAGTGTCCGAGGAGGGGAAAGTGGCTCTCTGCTGGGAGCTCTCT 600
154 ----- 154
601 CGGAGCGGGGAGAGCTCTCTCCCTGGCAGGAGACCTTCTCTCGCCAGTTCGATGGGCG 660
155 -----LysIleAsnLysLeuGluLysAlaValAlaAlaAlaHisThrPhePhe 170
661 ATGCTAACCCCAAGATCAACAAGTTGGAGAAGCTGTGTGTCGACGACACACCTTCTTC 720
171 ValGlyAsnProGluHisMetGluMetGlnGlnAsnLeuAspTyrTyrGlnThrMetSer 190
721 GTGGCAATCTCTGAGCATCATGGAATGCAGCAGAACCTAGACTATTACCAACCACTGCT 780
191 GlyValLysGluAlaAspPheLysAspLeuGluThrGlnProHisMetGlnGluPheArg 210
781 GGAGTGAAGAGGCGGACTTCAAGGATCTTCAAGACTCAACCCCATATGCAAGAAATTCGA 840
211 LeuGlyValArgLeuTyrSerGluGluGlnProGlnGluAlaValProHisLeuGluAla 230
841 CTGGGAGTGGGACTCTACTCAGAGGAACAGCCAGGAAAGCTGTGCCCCCAGCTAGAGCG 900
231 AlaLeuGlnGluTyrPheValAlaTyrGluGluCysArgAlaLeuCysGluGlyProTyr 250
901 GCGCTGCAAGATATCTTTGTGGCTATGAGAGTGGCGTGGCTCTCGGGAAGGCGCTAT 960
251 AspTyrAspGlyTyrAsnTyrLeuGluTyrAsnAlaAspLeuPheGlnAlaIleThrAsp 270
961 GACTACGATGGCTACACTACTCTTGTAGTACACGCTGACCTCTTCCAGGCCATCACAGAT 1020
271 HisTyrIleGlnValLeuAsnCysLysGlnAsnCysValThrGluLeuAlaSerHisPro 290
1021 CATTACATCCAGGTCTCACTGTAAAGCAAGAACTGTGTACGAGACTTGTCTCCACCACA 1080
291 SerArgGluLysProPheGluAspPheLeuProSerHisTyrAsnTyrLeuGlnPheAla 310
1081 AGTCGAGAGAAAGCCCTTTGAAGACTTCTCCCATCGCATTAATAATTATCTCAGTTTCC 1140
311 TyrTyrAsnIleGlyAsnTyrThrGlnAlaGlyCysAlaLysThrTyrLeuLeuPhe 330
1141 TACTATAACATGGGAATTATACACAGGCTGTGAATGTGCCAAGACCTATCTCTCTTC 1200
331 PheProAsnAspGluValMetAsnGlnAsnLeuAlaTyrTyrAlaAlaMetLeuGlyGlu 350
1201 TTCCCCCATGAGGAGTGATGAACCAAAATTTGGCTATTATGAGCTATGCTTTGGAGAA 1260
351 GluHisThrArgSerIleGlyProArgGluSerAlaLysGluTyrArgGlnArgSerLeu 370
```

Db 1261 GAACACACCATCCATCCGCGCCCGTGGAGTCCCAAGGAGTACCCAGACGGAAGCCTA 1320  
Qy 371 LeuGluLysGluLeuLeuPheAlaTyrAspValPheGlyIleProPheValAspPro 390  
Db 1321 CTGAAAAGAAAGTGCCTTTCTCGCTATGATGCTTTTGGAAATCCCTTTTGGATCCG 1380  
Qy 391 AspSerTrpThrProGluGluValIleProLysArgLeuGlnGluLysGlnLysSerGlu 410  
Db 1381 GATTTCAGTCCATCAGAAAGAGTGAATCCCAAGAGATTGCAAGAAACAGAAAGTCAGAA 1440  
Qy 411 ArgGluThrAlaValArgIleSerGlnGluIleGlyAsnLeuMetLysGluIleGluThr 430  
Db 1441 CGGAAAACAGCGCTGACGATCCCAAGAGATTGCGAAACCTTATGAAGGAATCGAGACC 1500  
Qy 431 LeuValGluGluLysThrLysGluSerLeuAspValSerArgLeuThrArgGluGlyGly 450  
Db 1501 CTTGTGAAGAGAGACCAAGAGTCACTGGATGTGAGCAGACTGACCCCGGAGGTGGC 1560  
Qy 451 ProLeuLeuTyrGluGlyIleSerLeuThrMetAsnSerLysLeuLeuAsnGlyTyrGln 470  
Db 1561 CCCCTGCTGTATGAAGCATCAGTCTCACCATGAATCCAAACTCCTGAAATGCTTCCAG 1620  
Qy 471 ArgValValMetAspGlyValIleSerAspHisGluCysGlnGluLeuGlnArgLeuThr 490  
Db 1621 CGGTGTGTGATGGACGGCGTAATCTTGACACAGAGTGTGAGGAGCTGCAGAGCTGACC 1680  
Qy 491 AsnValAlaAlaThrSerGlyAspGlyTyrArgGlyGlnThrSerProHisThrProAsn 510  
Db 1681 AATGTGCAGCAACCTCAGAGAGTGTCTACCGGGTCCAGACCTCCCCACATATCTCCCA 1740  
Qy 511 GluLysPheTyrGlyValThrValPheLysAlaLeuLysLeuGlyGlnGlyLysVal 530  
Db 1741 GAAAGATTCTATGTGTGCTACTGCTTCAAAGCCCTCAAGCTGGGGCAAGGCAAGATT 1800  
Qy 531 ProLeuGlnSerAlaHisLeuTyrTyrAsnValThrGluLysValArgAlaGlyMetGlu 550  
Db 1801 CCTCTGAGAGTGGCCACTGTACTACACGTGACGAGAGAGGTGGCGGCATCATGGAG 1860  
Qy 551 SerTyrPheArgLeuAspThrProLeuTyrPheSerTyrSerHisLeuValCysArgThr 570  
Db 1861 TCCTACTTCGCTCGGTGATAGCCCTCTACTTTCTACTCTCATCTGTTGTCGCGCACT 1920  
Qy 571 AlaIleGluGluValGlnAlaGluArgLysAspSerHisProValHisValAspAsn 590  
Db 1921 GCCATCGAAGAGGTCCAGGACGAGAGAGATGATGATCATCATGCTCCAGTGGACAC 1980  
Qy 591 CysIleLeuAsnAlaGluThrLeuValCysValLysGluProAlaTyrThrPheArg 610  
Db 1981 TGCATCTGTAATGCGAGACCTCGTGTGTCAAAGAGCCCGCAGCCTACACTTCGCG 2040  
Qy 611 AspTyrSerAlaIleLeuTyrLeuAsnGlyAspPheAspGlyGlyAsnPheTyrPheThr 630  
Db 2041 GACTACAGCGCCATCTCTTTTAACTTAAATGGGGACTTCGATGCGGAAACTTTTATTTC 2100  
Qy 631 GluLeuAspAlaLysThrValThrAlaGluValGlnProGlnCysGlyArgAlaValGly 650  
Db 2101 GAACCTGGATGCCAAGACCGTGACGGCAGAGGTGCAGCCTCAGTGTGAAGACCGTGG 2160  
Qy 651 PheSerSerGlyThrGluAsnProHisGlyValLysAlaValThrArgGlyGlnArgCys 670  
Db 2161 TTCTCTTCAGGACATGAAACCCACATGAGTGAAGGCTGTCAACGAGGGGAGCGCTGT 2220  
Qy 671 AlaIleAlaLeuTyrPheThrLeuAspProArgHisSerGluArgAspArgValGlnAla 690  
Db 2221 GCCATCGCCCTGTGTTTCACTTGGACCTTCGACACAGCGAGCGGACAGGGTGCAGG 2280  
Qy 691 AspAspLeuValLysMetLeuPheSerProGluGluMetAspLeuSerGlnGluGlnPro 710  
Db 2281 GATGACCTGGTGAAGATGCTCTTCAGCGCCAGAAAGATGAGACCTCTCCAGGAGAGCCC 2340  
Qy 711 LeuAspAlaGlnGlnGlyProProGluProAlaGlnGluSerLeuSerGlySerGluSer 730  
Db 2341 CTGGATGCCCCAGAGGGGCCCCCGAACCTTGCAACAGATCTCTCTCAGGCAATGATCG 2400

Qy 731 LysProLysAspGluLeu 736  
Db 2401 AAGCCCAAGGATGAGCTA 2418

## RESULT 6

US-09-728-952-51  
; Sequence 51, Application US/09728952  
; Patent No. US20020111302A1  
; GENERAL INFORMATION:  
; APPLICANT: Tang, Y. Tom  
; APPLICANT: Zhou, Ping  
; APPLICANT: Goodrich, Ryle  
; APPLICANT: Liu, Chenghua  
; APPLICANT: Asundi, Vinod  
; APPLICANT: Wang, Jian-Rui  
; APPLICANT: Wang, Dunrui  
; APPLICANT: Yamazaki, Vicki  
; APPLICANT: Ujwal, Manusha L.  
; APPLICANT: Drmanac, Radoje T.  
; TITLE OF INVENTION: No. US20020111302A1el Nucleic Acids and  
; TITLE OF INVENTION: Polypeptides  
; FILE REFERENCE: 799  
; CURRENT APPLICATION NUMBER: US/09/728,952  
; CURRENT FILING DATE: 2000-11-30  
; NUMBER OF SEQ ID NOS: 101  
; SOFTWARE: pt\_FL\_genes Version 2.0  
; SEQ ID NO 51  
; LENGTH: 2753  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
; FEATURE:  
; NAME/KEY: CDS  
; LOCATION: (1)..(2418)  
US-09-728-952-51

Alignment Scores:  
Pred. No.: 0 Length: 2753  
Score: 3801.00 Matches: 733  
Percent Similarity: 90.94% Conservative: 0  
Best Local Similarity: 90.94% Mismatches: 3  
Query Match: 98.22% Indels: 70  
DB: 9 Gaps: 1

US-10-045-815-4 (1-736) x US-09-728-952-51 (1-2753)

Qy 1 MetAlaValArgAlaLeuLysLeuLeuThrThrLeuLeuAlaValAlaAlaSer 20  
Db 1 ATGGCGGTACGCGCTTGAAGCTGTGACACACACTGCTGGCTGTGCGGCGCTGCTCC 60  
Qy 21 GlnAlaGluValGluSerGluAlaGlyTyrGlyMetValThrProAspLeuLeuPheAla 40  
Db 61 CAAGCCGAGGTGAGTCCGAGCGAGGATGGGCGATGGAGCGCTGATCTCTCTTCGCC 120  
Qy 41 GluGlyThrAlaAlaTyrAlaArgGlyAspTrpProGlyValValLeuSerMetGluArg 60  
Db 121 GAGGGAGCGCGAGCCTACGCGCGGGGACTGGCCCGGGTGGTCTCTGAGCATGGAACGG 180  
Qy 61 AlaLeuArgSerArgAlaAlaLeuArgAlaLeuArgLeuArgCysArgThrGlnCysAla 80  
Db 181 GCGCTGCGCTCCCGGGCAGCCCTCCGCGCCCTTCCGCTGCGCTCCGCGCACCCAGTGTGCC 240  
Qy 81 AlaAspPheProTyrGluLeuAspProAspTrpSerProSerProAlaGlnAlaSerGly 100  
Db 241 GCGGACTTCCCTGGAGCTGAGACCCGAGCTGGTCCCGCCCGCGCCAGGCTCGGGC 300  
Qy 101 AlaGlyAlaLeuArgAspLeuSerPhePheGlyGlyLeuLeuArgArgAlaAlaCysLeu 120  
Db 301 GCGCGCGCCCTGCGGACCTGAGCTTCTTCGGGGGCGCTTTCGCGTCCGCTGCGCTGCTG 360  
Qy 121 ArgArgCysLeuGlyProProAlaAlaHisSerLeuSerGluGluMetGluLeuGluPhe 140  
Db 361 CGCGCTGCTCCGCGCGCGCGCGCCCACTCGCTCAGCGAAGAGATCGACCTGGAGTTC 420

141 ArgLysArgSerProTyrAsnTyrLeuGlnValAlaTyrPhe----- 154  
 Db CGCAAGCGGAGCCCTACAACTACCTGCAGGTCCGCTTCTTCAAGGTGCAGACCTGCCTG 480  
 QY 154 ----- 154  
 Db 481 GAACAGCGCGCGGGTCTCTCTGGGAGAGAGTGTTCAGGGGACCTGAGGAGCTTG 540  
 QY 154 ----- 154  
 Db 541 GGGGATCGGGAAGTGTCCGAGGAGGGAAGTGGCTCTCTGGCTGGGGAGCTCTCCT 600  
 QY 154 ----- 154  
 Db 601 CGGAGCGCGGAGAGCTGCTCTCTGGCAGGAGACCTTCTCTCGCCCGAGTTGCGATGGGCGAG 660  
 QY 155 -----LysIleAsnLysLeuGluLysAlaValAlaAlaAlaHisThrPhePhe 170  
 Db 661 ATGCTAACCCCAAGATCAACAGATTGGAGAAAGCTGTCTCGACGACACACCTTCTTTC 720  
 QY 171 ValGlyAsnProGluHisMetGluMetGlnGlnAsnLeuAspTyrTyrGlnThrMetSer 190  
 Db 721 GTGGCAATCTCGAGCATGTGAAATGCAGCAGAACCTAGACTATTACCAACCATGTCT 780  
 QY 191 GlyValLysGluAlaAspPheLysAspLeuGluThrGlnProHisMetGlnGluPheArg 210  
 Db 781 GGAGTGAAGGAGCGGACTTCAAGGATCTTGAGACTCAACCCCATATGCAAGAAATTCGA 840  
 QY 211 LeuGlyValArgLeuTyrSerGluGluGlnProGlnGluAlaValProHisLeuGluAla 230  
 Db 841 CTGGGAGTGCAGCTTACTACAGAGAACACCCACAGAGCTGTGCCACCTAGAGGCG 900  
 QY 231 AlaLeuGlnGluTyrPheValAlaTyrGluGluCysArgAlaLeuCysGluGlyProTyr 250  
 Db 901 CGCGTGAAGAATACCTTGTGGCTATGAGGAGTGGCGTCCCTCTCGCAAGGCGCCCTAT 960  
 QY 251 AspTyrAspGlyTyrAsnTyrLeuGluTyrAsnAlaAspLeuPheGlnAlaIleThrAsp 270  
 Db 961 GACTACAGTGGCTACAACTACCTTGAGTACAGCTGACCTTTCAGGCGCATACAGAT 1020  
 QY 271 HisTyrIleGlnValLeuAsnCysLysGlnAsnCysValThrGluLeuAlaSerHisPro 290  
 Db 1021 CATTATACAGGTCTCAACTGTAAAGCAGAACTGTGTACGAGCTTGTCTCCACCCA 1080  
 QY 291 SerArgLysProPheGluAspPheLeuProSerHisTyrAsnTyrLeuGlnPheAla 310  
 Db 1081 AGTCGAGAGAGCCCTTTGAAGACTTCTCTCCATCGCATTTATATATCTGCGAGTTGCC 1140  
 QY 311 TyrTyrAsnIleGlyAsnTyrThrGlnAlaGlyGluCysAlaLysThrTyrLeuLeuPhe 330  
 Db 1141 TACTATAACATTGGGAATTATACACAGCTGTTGAATGTCGAAGACCTATCTCTCTTC 1200  
 QY 331 PheProAsnAspGluValMetAsnGlnAsnLeuAlaTyrTyrAlaAlaMetLeuGlyGlu 350  
 Db 1201 TTCCCCCAATGACGAGGTGATGAACCAAAATTTGGCCCTATTATGAGCTATGCTGGAGAA 1260  
 QY 351 GluHisThrArgSerIleGlyProArgGluSerAlaLysGluTyrArgGlnArgSerLeu 370  
 Db 1261 GAACACACCATCATCGGCCCTCGTGAAGTGCCAAAGAGTACCGACAGCGAGCCTA 1320  
 QY 371 LeuGluLysGluLeuLeuPhePheAlaTyrAspValPheGlyIleProPheValAspPro 390  
 Db 1321 CTGGAAAAGAACCTGCTTCTCGCTATGATGTTTTTGGAAATTCCTTTGTGGATCCG 1380  
 QY 391 AspSerThrProGluGluValIleProLysArgLeuGlnGluLysGlnLysSerGlu 410  
 Db 1381 GATTTCATGGATCCAGAAAGATGATTTCCCAAGAGATTGCAAGAGAAACAGAAAGTCAGAA 1440  
 QY 411 ArgGluThrAlaValArgIleSerGlnGluIleGlyAsnLeuMetLysGluIleGluThr 430  
 Db 1441 CGGSAACACGCGTACGCATCTCCACGAGATTGGGAACCTTATGAAGGAAATCGAGACC 1500

QY 431 LeuValGluGluLysThrLysGluSerLeuAspValSerArgLeuThrArgGluGlyGly 450  
 Db 1501 CTTTGTGGAAGAGAACCAAGAGTCACTGGATGTGAGCAGACTGACCCGGGAAGGTGGC 1560  
 QY 451 ProLeuLeuTyrGluGlyLysSerLeuThrMetSerLysLeuLeuAsnGlyTyrGln 470  
 Db 1561 CCCCTGCTGTATGAAGGCATCAGTCTCACCATGAATCCAAACTCCTGAATGTTCCCG 1620  
 QY 471 ArgValValMetAspGlyValIleSerAspHisGluCysGlnGluLeuGlnArgLeuThr 490  
 Db 1621 CGGGTGGTGATGAGCGCGTAACTCTCTCACCAGTGTCTCAGGAGCTGCAGAGACTGACC 1680  
 QY 491 AsnValAlaAlaThrSerGlyAspGlyTyrArgGlyGlnThrSerProHisThrProAsn 510  
 Db 1681 AATGTGGCAGCAACCTCAGGAGATGGCTACCGGGGTGAGACTCCCCACATCTCCCAAT 1740  
 QY 511 GluLysPheTyrGlyValThrValPheLysAlaLeuLysLeuGlyGlnGlyLysVal 530  
 Db 1741 GAAAAGTTCTATGGTGTCTACTGTCTTCAAGCCCTCAAGCTGGGCAAGAGGCAAGT 1800  
 QY 531 ProLeuGlnSerAlaHisLeuTyrTyrAsnValThrGluLysValArgArgIleMetGlu 550  
 Db 1801 CCTCTGCAGAGTGGCCACCTGTACTACACGTCAGGAGAGAGGTGCGGCGCATCATGAG 1860  
 QY 551 SerTyrPheArgLeuAspThrProLeuTyrPheSerTyrSerHisLeuValCysArgThr 570  
 Db 1861 TCTTACTTCCGCTGGATACGCCCCCTCTACTTTTCTTACTCTCATCTGCTGCGGCACT 1920  
 QY 571 AlaIleGluGluValGlnAlaGluLysAspSerHisProValHisValAspAsn 590  
 Db 1921 GCCATCGAAGAGTCCAGGCGAGAGAGGATGATAGTATCATCATCAGTCCAGTGACAA 1980  
 QY 591 CysIleLeuAsnAlaGluThrLeuValCysValLysGluProProAlaTyrThrPheArg 610  
 Db 1981 TGCACTCTGAAATGCGAGACCTCGTGTGTGTCAAGAGCGCCCGCTACACCTTCGCG 2040  
 QY 611 AspTyrSerAlaIleLeuTyrLeuAsnGlyAspPheAspGlyGlyAsnPheTyrPheThr 630  
 Db 2041 GACTACAGCGCATCTTTACTTAAATGGGACCTTCGATGGGGAACCTTTTATTTCACT 2100  
 QY 631 GluLeuAspAlaLysThrValThrAlaGluValGlnProGlnCysGlyArgAlaValGly 650  
 Db 2101 GAACTGGATGTCGAAGCCGTGACGCGAGAGTGCAGCTCAGCTGTGGAAGAGCGGTGG 2160  
 QY 651 PheSerSerGlyThrGluAsnProHisGlyValLysAlaValThrArgGlyGlnArgCys 670  
 Db 2161 TTCTCTTTCAGGACCTGAAACCCACATGAGTGAAGGTGTTCACCGAGGGGCGAGCTGT 2220  
 QY 671 AlaIleAlaLeuThrPheThrLeuAspProArgHisSerGluArgAspArgValGlnAla 690  
 Db 2221 GCCATCGCCCTGTGCTTTCACCTGGACCTCGACACAGCAGCGGAGCAGGAGTGCAGG 2280  
 QY 691 AspAspLeuValLysMetLeuPheSerProGluGluMetAspLeuSerGlnGluGlnPro 710  
 Db 2281 GATGACCTGGTGAAGATGCTCTTCAGCCCAAGAGATGGACCTCTCCAGAGAGCGGCC 2340  
 QY 711 LeuAspAlaGlnGlnGlyProProGluProAlaGlnGluSerLeuSerGlySerGluSer 730  
 Db 2341 CTGGATGCCAGCAGCGGCCCTCCCGAACCTGCACAAAGAGTCTCTCTCAGGAGCTGAATCG 2400  
 QY 731 LysProLysAspGluLeu 736  
 Db 2401 AAGCCCAAGGATGAGCTA 2418

RESULT 7  
 US-10-045-815-1  
 ; Sequence 1, Application US/10045815  
 ; Publication No. US20020160498A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Wadhwa, Renu  
 ; APPLICANT: Sugihara, Takashi  
 ; APPLICANT: Ohide, Akiko  
 ; TITLE OF INVENTION: TUMOR SUPPRESSOR GENE

```
FILE REFERENCE: 06501-091001
CURRENT APPLICATION NUMBER: US/10/045.815
CURRENT FILING DATE: 2001-10-26
PRIOR APPLICATION NUMBER: PCT/JP00/02731
PRIOR FILING DATE: 2000-04-26
PRIOR APPLICATION NUMBER: JP 11/118806
PRIOR FILING DATE: 1999-04-26
NUMBER OF SEQ ID NOS: 15
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 1
LENGTH: 2829
TYPE: DNA
ORGANISM: Homo sapiens
FEATURE:
NAME/KEY: CDS
LOCATION: (52)...(1140)
US-10-045-815-1

Alignment Scores:
Pred. No.: 0 Length: 2829
Score: 3780.00 Matches: 733
Percent Similarity: 90.41% Conservative: 2
Best Local Similarity: 90.16% Mismatches: 1
Query Match: 97.67% Indels: 77
DB: 14 Gaps: 2

US-10-045-815-4 (1-736) x US-10-045-815-1 (1-2829)

Qy 1 MetAlaValArgAlaLeuLeuLysLeuLeuThrThrLeuLeuAlaValAlaAlaSer 20
Db 52 ATGCGGTACCGCGGTGAAGCTGCTGACACACTGCTGCTGCTGCTGCTGCTGCTGCT 111
Qy 21 GlnAlaGluValGluSerGluAlaGlyTrpGlyMetValThrProAspLeuLeuPheAla 40
Db 112 CAAGCCGAGTTCGAGTCGAGGAGGATGGGATGGTGTGCTGCTGCTGCTGCTGCTGCT 171
Qy 41 GluGlyThrAlaAlaValArgAlaGlyAspTrpProGlyValValLeuSerMetGluArg 60
Db 172 GAGGGACCGCAGCTACCGCGCGCGGAGCTGGCGCGGCTGCTGCTGCTGCTGCTGCTGCT 231
Qy 61 AlaLeuArgSerArgAlaAlaLeuArgAlaLeuArgLeuArgCysArgThrGlnCysAla 80
Db 232 GCGCTGCGCTCCGCGGAGCGCTCCGCGCGCTTCCGCTGCGCTGCGCGCACCCAGTGTGCC 291
Qy 81 AlaAspPheProTrpGluLeuAspProAspTrpSerProSerProAlaGlnAlaSerGly 100
Db 292 GCCGACTTCGCGTGGAGCTGGACCCCGACCTGGTTCCTCCCGCGCGCGCTCGGGC 351
Qy 101 AlaGlyAlaLeuArgAspLeuSerPhePheGlyGlyLeuLeuArgArgAlaAlaCysLeu 120
Db 352 GCCGCGCGCTTCGCGCGCTGAGCTTCTTCGCGCGCGCTTCTGCGTTCGCGCTGCTGCTG 411
Qy 121 ArgArgCysLeuGlyProProAlaAlaHisSerLeuSerGluGluMetGluLeuPhe 140
Db 412 CGCCGCTGCTTCGCGCGCGCGCGCGCTCCGCTCAGCGAGAGATGGAGCTGGAGTTC 471
Qy 141 ArgLysArgSerProTrpArgTrpLeuGlnValAlaTrpPheLysLeuLeuLysLeuGlu 160
Db 472 CGCAAGCGGAGCGCTCAACTACTCGCTGAGTTCGCTGAGTTCGCTGAGTTCGCTGAGTTC 531
Qy 161 LysAlaValAlaAlaAlaHisThrPhePheValGlyAsnProGluHisMetGluMetGln 180
Db 532 AAAGCTGTGCTGAGCAGACACCTTCTTCGTTGGGCAATCTGAGCAGATGGAAATGCGAG 591
Qy 181 GlnAsnLeuAspTrpTrpGlnThrMetSerGlyValLysGluAlaAspPheLysAspLeu 200
Db 592 CAGAACCTAGACTATTACCAAAACCATGCTGAGTGAAGGAGCGGAGCTTCAAGATCAACAA 651
Qy 201 GluThrGlnProHisMetGlnGluPheArgLeuGlyValArgLeuTrpSerGluGln 220
Db 652 GAGACTCAACCCCATATGCAAGAAATTTGCTGAGTGGAGTGGAGTGGAGTGGAGTGGAG 711
Qy 221 ProGlnGluAlaValProHisLeuGluAlaAlaLeuGlnGluTrpPheValAlaTrpGlu 240

712 CCACAGGAAGCTGTGCTCCCACTAGAGCGCGCTGCAAGAAATATTGTGGCTATGAG 771
241 GluCysArgAlaLeuCysGluGlyProTrpArgTrpAspGlyTrpAsnTrpLeuGluTrp 260
772 GAGTCCGCTGCTCTCGAAGGCGCTTATGATGATGATGATGATGATGATGATGATGATGAT 831
261 AsnAlaAspLeuPheGlnAlaLeuThrAspHisTrpLeuGlnValLeuAsnCysLysGln 280
832 AACGCTGACCTCTTCAGGCGCATCATCATCATCATCATCATCATCATCATCATCATCATCAT 891
281 AsnCysValThrGluLeuAlaSerHisProSerArgGluLysProPheGluAspPheLeu 300
892 AAATGTGTGTCAGGAGCTTGTCTTCCCAAGTCAGAGAGAGAGAGAGAGAGAGAGAGAGAG 951
301 ProSerHisTrpAsnTrpLeuGlnPheAlaTrpTrpAsnTrpLeuGlnAlaAla 320
952 CCATCGCATTAATAATTATCTGCACTTTCCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCT 1011
321 GlyGluCysAlaLysThrTrpLeuLeuPhePheProAsnAspGluValMetAsnGlnAsn 340
1012 GGTGAATGTCCAGACCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCT 1071
341 LeuAlaTrpTrpAlaAlaMetLeuGlyGluGluHisThrArgSerIleGlyProArgGlu 360
1072 TTGGCTATTATGCACTATGCTTGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1131
361 -----SerAl 362
1132 CAGGCGCACTAGGAAAGATGTGACCCCGGAAAGTACTCAGTTTCCCTGCCCTCGAGTGC 1191
362 alyGluTrpArgGlnArgSerLeuLeuGluLysLeuLeuLeuPheAlaTrpAspVa 382
1192 CAAGAGTACCGACAGCGAGCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCT 1251
382 lPheGlyLeuProPheValAspProAspSerTrpThrProGluGluValIleProLysAr 402
1252 TTTTGGAAATCCCTTTGTGATCGGATTCATGACCTCCAGAGAGAGAGAGAGAGAGAGAGAG 1311
402 gLeuGlnGluLysGlnLys----- 408
1312 ATTGCAAGAGAAACAGAGAGTGGAGACCTTGAAGAACTGTCATGTTGGATCAGTCTGATG 1371
408 ----- 408
1372 AAGCACTTGAGGCTTCTTGTGAGCCAGGAGAGATGTGAACCTCTGCGAAGGGGTGGCAGGT 1431
408 ----- 408
1432 CCAGTTTGGGAAGCTCGGGGTGGAGCCCGAGGCTGGCCCTCGAATGAGTCTCTAGAGCGG 1491
409 -----SerGluArgGluThrAlaValArgIleSerGlnGluIleGlyAsnLe 424
1492 TTGTGCTCATAGTTCAGAAACGGGAAACAGCGCTACGATCTCCAGGAGATGGGACCT 1551
424 uMetLysGluLeuGluThrLeuValGluGluLysThrLysGluSerLeuAspValSerAr 444
1552 TATGAAGAAATCGAGACCTTTGTGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1611
444 gLeuThrArgGluGlyGlyProLeuLeuTrpGluGlyIleSerLeuThrMetAsnSerLy 464
1612 ACTGACCCGGGAAGGTGGCCCCCTGCTGTATGAAGGATCAGTCTCACCAGTACTCAA 1671
464 sLeuLeuAsnGlyTrpGlnArgValMetAspGlyValIleSerAspHisGluCysGl 484
1672 ACTCTGAATGGTTACCAAGCGGCTGGTATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1731
484 nGluLeuGlnArgLeuThrAsnValAlaAlaThrSerGlyAspGlyTrpArgGlyGlnTh 504
1732 GGAGCTGAGAGACTGACCAATGTGGCAGCAGCACTCAGGAGATGGCTACCGGGTCCAGAC 1791
504 rSerProHisThrProAsnGluLysPheTrpGlyValThrValPheLysAlaLeuLysLe 524
```



Db 1792 CTCCACATACCTCCCAATGAAGAAGTTCTATGGTGTCACTGTCTTCAAAGCCCTCAAGCT 1851  
Qy 524 uGlyGlnGluGlyValProLeuGlnSerAlaHisLeuTyrTyrAsnValThrGluGly 544  
Db 1852 GGGCCAAAGAGGAAAGTTCTCTGAGAGTGCACCTGTACTACAACGTCAGCAGAGAA 1911  
Qy 544 sValArgArgIleMetGluSerTyrPheArgLeuAspThrProLeuTyrPheSerTyrSe 564  
Db 1912 AGTGGCGGCATCATGGAGTCTTCCGCTGGATACGCCCTCTACTTTCTTCTACTC 1971  
Qy 564 rHisLeuValCysArgThrAlaIleGluGlnValGlnAlaGluArgLysAspAspSerHi 584  
Db 1972 TCATCTGGTGTGCGGCACTGCCATCGAAGAGGTCCAGGCAGAGAGGAAAGCATAGTCA 2031  
Qy 584 sProValHisValAspAsnCysTleLeuAsnAlaGluThrLeuValCysValLysLeuPr 604  
Db 2032 TCCAGTCCACGTCGACCACTGCACTCTGAATGCCGAGACCTCGTGTGTGTCAAGAGCC 2091  
Qy 604 oProAlaTyrThrPheArgAspTyrSerAlaIleLeuTyrLeuAsnGlyAspPheAspGl 624  
Db 2092 CCAGAGCTACACCTTCGGGACTACAGCGCCATCCTTTACCTAAATGGGACTTCGATGG 2151  
Qy 624 yGlyAsnPheTyrPheThrGluLeuAspAlaLysThrValThrAlaGluValGlnProGl 644  
Db 2152 CGAAACCTTTTATTTCACTGAATGGATGCCAAGACCGTGACGGCAGAGGTGCAGCCTCA 2211  
Qy 644 nCysGlyArgAlaValGlyPheSerSerGlyThrGluAsnProHisGlyValLysAlaVa 664  
Db 2212 GTGTGAAGAGCCGCTGGATTCCTTCAGGCACTGAACCCACATGGAGTGAAGCTGT 2271  
Qy 664 lThrArgGlyGlnArgCysAlaIleAlaLeuTyrPheThrLeuAspProArgHisSerGl 684  
Db 2272 CACCGGGGGCAGCGCTGTGCCATCGCCTGTGGTTCACTGGACCCCTCGACACACGCA 2331  
Qy 684 uArgAspArgValGlnAlaAspAspLeuValLysMetLeuPheSerProGluMetAs 704  
Db 2332 GCGGACAGGAGGTGACGAGATGACCTGGTGAAGATGCTCTTCAGGCCCAAGAGATGGA 2391  
Qy 704 pLeuSerGlnGlnProLeuAspAlaGlnGlnGlyProProGluProAlaGlnGluSe 724  
Db 2392 CCTCTCCAGGAGAGCCCTGGATGCCAGCAGAGGGCCCCCGAACCTCGACAGAGTC 2451  
Qy 724 rLeuSerGlySerGluSerLysProLysAspGluLeu 736  
Db 2452 TCTCTCAGGCAGTGAATCGAAGCCCAAGGATGAGCTA 2488

RESULT 8  
US-10-302-172-227  
; Sequence 227, Application US/10302172  
; Publication No. US20040053250A1  
; GENERAL INFORMATION:  
; APPLICANT: Tang, Y. Tom  
; APPLICANT: Xue, Aidong J.  
; APPLICANT: Dumanac, Radoje T.  
; TITLE OF INVENTION: No. US20040053250A1el Arginine-rich Protein-like Nucleic Acids an  
; TITLE OF INVENTION: Polypeptides  
; FILE REFERENCE: 803\_ICNCP  
; CURRENT APPLICATION NUMBER: US/10/302,172  
; CURRENT FILING DATE: 2002-11-21  
; PRIOR APPLICATION NUMBER: US 10/225,251  
; PRIOR FILING DATE: 2002-08-20  
; PRIOR APPLICATION NUMBER: PCT US02/05095  
; PRIOR FILING DATE: 2002-03-05  
; PRIOR APPLICATION NUMBER: US 09/799,451  
; PRIOR FILING DATE: 2001-03-05  
; NUMBER OF SEQ ID NOS: 950  
; SOFTWARE: pt FL\_genes Version 2.0  
; SEQ ID NO 227  
; LENGTH: 2152  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
; FEATURE:  
; NAME/KEY: CDS

; LOCATION: (32)...(2149)  
US-10-302-172-227  
Alignment Scores:  
Pred. No.: 0 Length: 2152  
Score: 3645.00 Matches: 699  
Percent Similarity: 95.24% Conservative: 2  
Best Local Similarity: 94.97% Mismatches: 5  
Query Match: 94.19% Indels: 30  
DB: 13 Gaps: 1

US-10-045-815-4 (1-736) x US-10-302-172-227 (1-2152)

Qy .1 MetAlaValArgAlaLeuLysLeuThrThrLeuAlaValValAlaAlaAser 20  
Db 32 ATGGCGTACCGCGGTGAAGCTGTGACACACACTGTGGCTGTCTGGCCGCTGCCTCC 91  
Qy 21 GlnAlaGluValGluSerGluAlaGlyTyrGlyMetValThrProAspLeuLeuPheAla 40  
Db 92 CAAGCCGAGGTGCGATCCGAGCGAGGATGGCGCATGTGTGACGCTGATCTCTCTTCC 151  
Qy 41 GluGlyThrAlaAlaTyrAlaArgGlyAspTyrProGlyValValLeuSerMetGluArg 60  
Db 152 GAGGGACCGCAGCTACCGCGCGGAGCTGGCCGGGTGGTCTCTGAGCATGGACGG 211  
Qy 61 AlaLeuArgSerArgAlaAlaLeuArgAlaLeuArgLeuArgCysArgThrGlnCysAla 80  
Db 212 GCGCTGCGCTCCCGCGCAGCCCTCCGCGCCCTTCGCTGCGCTCCGCGCAGCTGTGCC 271  
Qy 81 AlaAspPheProTyrGluLeuAspProAspTyrPheSerProAlaGlnAlaSerGly 100  
Db 272 GCGGATCTCCGCGGAGCTGGACCCCGACCTGGTGTCCCGCCAGCCCGCCAGGCTTCGG 331  
Qy 101 AlaGlyAlaLeuArgAspLeuSerPheGlyGlyLeuLeuArgArgAlaAlaCysLeu 120  
Db 332 GCGCGCGCTCCGCGACCTGAGCTTCTTCGGGGGCTCTTCGCTGCGCTGCGCTGCTG 391  
Qy 121 ArgArgCysLeuGlyProProAlaAlaHisSerLeuSerGluGluMetGluLeuGluPhe 140  
Db 392 GCGCGCTCCCTCGGCGCCCGCCGCTCGCTCGTCAAGAGAGATGAGCTGAGTTC 451  
Qy 141 ArgLysArgSerProTyrAsnTyrLeuGlnValAlaTyrPheLysIleAsnLysLeuGlu 160  
Db 452 CGCAAGCGAGCCCTCAACTACTCTGCGAGTTCGCTACTTCAAGATCAACAGTTGGAG 511  
Qy 161 LysAlaValAlaAlaAlaHisThrPheValGlyAsnProGluHisMetGluMetGln 180  
Db 512 AAAGCTGTGTGCGACGACACACCTTCTTCGTTGGGCAATCTCGAGCATGGAATGCGAG 571  
Qy 181 GlnAsnLeuAspTyrTyrGlnThrMetSerGlyValLysGluAlaAspPheLysAspLeu 200  
Db 572 CAGAACCTTAGACTATTACCAACCATGTCTGAGTGAAGAGGCGGCTTCAAGATCTT 631  
Qy 201 GluThrGlnProHisMetGlnGluPheArgLeuGlyValArgLeuTyrSerGluGln 220  
Db 632 GAGACTCAACCCCATATCAAGAAATTTTCGACTGGGAGTGGACTCTACTCAGAGGAACAG 691  
Qy 221 ProGlnGluAlaValProHisLeuGluAlaLeuGlnGluTyrPheValAlaTyrGlu 240  
Db 692 CCACAGGAAGTGTGCGCCACCTAGAGCGCGCTGCAAGAAATACCTTTGTGGCTTATGAG 751  
Qy 241 GluCysArgAlaLeuCysGluGlyProTyrAspTyrAspGlyTyrAsnTyrLeuGluTyr 260  
Db 752 GAGTCCGCTGCGCTCTGGAAGGGCCCTATGACTACGATGGCTACAACTACCTTGGATC 811  
Qy 261 AsnAlaAspLeuPheGlnAlaIleThrAspHisTyrIleGlnValLeuAsnCysLysGln 280  
Db 812 AACGCTGACCTTCTCCAGGCCATCACAGATCATTAATCCAGGTCTCTCACTGTAAAGAG 871  
Qy 281 AsnCysValThrGluLeuAlaSerHisProSerArgGluLysProPheGluAspPheLeu 300  
Db 872 AACTGTGTACGGAGCTTGTCTTCCACCCCAAGTTCGAGAGAGCCCTTTGAAGACTTCTCTC 931

QY 301 ProSerHisTyrAsnTyrLeuGlnPheAlaTyrTyrAsnIleGlyAsnTyrThrGlnAla 320  
DB 932 CCATCGCAATTAATAATCTCGAGTTGGCTTACTATACATTTGGGAATTATACACAGGCT 991  
QY 321 GlyGluCysAlaLysThrTyrLeuLeuPhePheProAsnAspGluValMetAsnGlnAsn 340  
DB 992 GTTGAATGTCGAAGACCTATCTCTCTCTTCCCAATGACGAGGTGATGAACCAAAAT 1051  
QY 341 LeuAlaTyrTyrAlaAlaMetLeuGlyGluGluHisThrArgSerIleGlyProArgGlu 360  
DB 1052 TTGGCTTATTATGAGCTATGCTTGGAGAGAACACACACATCCATCGGCCCCCGTGAG 1111  
QY 361 SerAlaLysGluTyrArgGlnArgSerLeuLeuGluLysGluLeuLeuPheAlaTyr 380  
DB 1111 ----- 1111  
QY 391 AspValPheGlyIleProPheValAspProAspSerTrpThrProGluGluValIlePro 400  
DB 1112 -----GATTTCATGGACTCCAGGAAGTGAATTCCTCC 1141  
QY 401 LysArgLeuGlnGluLysGlnLysSerGluArgGluThrAlaValArgIleSerGlnGlu 420  
DB 1142 AAGAGATTGCAAGAAACAGAAATGACAGCGGAACAGCCGACCACTCTCCAGAG 1201  
QY 421 IleGlyAsnLeuMetLysGluIleGluThrLeuValGluGluLysThrLysGluSerLeu 440  
DB 1202 ATTGGGAACCTTATGAAGAAATCGAGACCTTGTGAAGAGAAGCAAGGAGTCACTG 1261  
QY 441 AspValSerArgLeuThrArgGluGlyGlyProLeuLeuTyrGluGlyIleSerLeuThr 460  
DB 1262 GATGTGACAGACTGACCCGGAGAGTGGCCCTCTGCTGATGAAGGCATTAGTCTCACC 1321  
QY 461 MetAsnSerLysLeuLeuAsnGlyTyrGlnArgValValMetAspGlyValIleSerAsp 480  
DB 1322 ATGAATCCAAACTCTTGTGTTGCCAGGGTGTGATGGCGCGTAATCTCTGAC 1381  
QY 481 HisGluCysGlnGluLeuGlnArgLeuThrAsnValAlaAlaThrSerGlyAspGlyTyr 500  
DB 1382 CACCAATGTACAGAGCTGCAGAGACTGACCAATGGGCGAGCAACTCAGGAGATGGCTAC 1441  
QY 501 ArgGlyGlnThrSerProHisThrProAsnGluLysPheTyrGlyValThrValPheLys 520  
DB 1442 CGGGGTGAGACTCCCACTATCTCCCAATGAAAGTTCTATGGTGTCACTGCTTCAAA 1501  
QY 521 AlaLeuLysLeuGlyGlnGluGlyLysValProLeuGlnSerAlaHisLeuTyrTyrAsn 540  
DB 1502 GCCCTCAAGCTGGGGCAAGAGGCAAAAGTTCCTTCGACAGAGTGCCCACTGTACTACAAC 1561  
QY 541 ValThrGluLysValArgArgIleMetGluSerTyrPheArgLeuAspThrProLeuTyr 560  
DB 1562 GTGACGAGAAAGTGGCGCGCATCATGGAGTCTTCTCCCTGGATAGCCCTCTAC 1621  
QY 561 PheSerTyrSerHisLeuValCysArgThrAlaIleGluValGlnAlaGluArgLys 580  
DB 1622 TTTTCTACTCTCATCTGGTGTGGCGACTGCCATCGAAGAGTCCAGGCGAGAGAGGAAG 1681  
QY 581 AspAspSerHisProValHisValAspAsnCysIleLeuAsnAlaGluThrLeuValCys 600  
DB 1682 GATGATAGTATCCAGTCCAGTGGCAACTGATCTCTGAATCCGAGAGACCTCGTGTGT 1741  
QY 601 ValLysGluProProAlaTyrThrPheArgAspTyrSerAlaIleLeuTyrLeuAsnGly 620  
DB 1742 GTCAAGAGAGCCCTCAGCTACCTCCGCGACTACAGCGCATCTTTTACCTAAATGGG 1801  
QY 621 AspPheAspGlyGlyAsnPheThrPheThrGluLeuAspAlaLysThrValThrAlaGlu 640  
DB 1802 GACTTCGATGGCGAAATCTTTATTTTCATCTGAATCCCAAGACCGTGACGGCAGAG 1861  
QY 641 ValGlnProGlnCysGlyArgAlaValGlyPheSerSerGlyThrGluAsnProHisGly 660  
DB 1862 GTGCAGCTCAGTGTGAAGAGCGGTGGATTCTCTTCAGGCACTGAAACCCACATGGA 1921  
QY 661 ValLysAlaValThrArgGlyGlnArgCysAlaIleAlaLeuThrPheThrLeuAspPro 680

DB 1922 GTGAGAGCTGTCCACAGGGGGCAGCGCTGTGCATCGCCCTGTGTTTCCCTTGACCT 1981  
QY 681 ArgHisSerGluArgAspArgValGlnAlaAspAspLeuValLysMetLeuPheSerPro 700  
DB 1982 CGACACAGCGAGCGGACAGGGGTGAGGACATGACCTGTGGTGAAGATGCTCTTACGCCA 2041  
QY 701 GluGluMetAspLeuSerGlnGlnGlnProLeuAspAlaGlnGlnGlyProGluPro 720  
DB 2042 GAAGAGATGGACTCTCCAGGAGCAGCCCTGGATGCCAGAGGGCCCCCGAACCT 2101  
QY 721 AlaGlnGluSerLeuSerGlySerGluSerLysProLysAspGluLeu 736  
DB 2102 GCACAAGAGTCTCTCTCAGGCACTGAATCGAAGCCCAAGGATGAGCTA 2149  
RESULT 9  
US-10-257-174-1  
; Sequence 1, Application US/10257174  
; Publication No. US20040034194A1  
; GENERAL INFORMATION:  
; APPLICANT: Agarwal, Pankaj  
; APPLICANT: Murdoch, Paul R.  
; APPLICANT: Rizvi, Safia K.  
; APPLICANT: Smith, Randall F.  
; APPLICANT: Xiang, Zhaoying  
; TITLE OF INVENTION: NOVEL COMPOUNDS  
; FILE REFERENCE: GP50022  
; CURRENT APPLICATION NUMBER: US/10/257,174  
; CURRENT FILING DATE: 2002-10-10  
; PRIOR APPLICATION NUMBER: PCI/US01/11797  
; PRIOR FILING DATE: 2001-04-11  
; PRIOR APPLICATION NUMBER: 60/196,603  
; PRIOR FILING DATE: 2000-04-13  
; PRIOR APPLICATION NUMBER: 60/199,417  
; PRIOR FILING DATE: 2000-04-24  
; NUMBER OF SEQ ID NOS: 48  
; SOFTWARE: FastSeq for Windows Version 3.0  
; SEQ ID NO 1  
; LENGTH: 2127  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
US-10-257-174-1  
Alignment Scores:  
Pred. No.: 0 Length: 2127  
Score: 3572.00 Matches: 691  
Percent Similarity: 94.57% Conservative: 5  
Best Local Similarity: 93.89% Mismatches: 12  
Query Match: 92.30% Indels: 28  
DB: 13 Gaps: 3  
US-10-045-815-4 (1-736) x US-10-257-174-1 (1-2127)  
QY 1 MetAlaValArgAlaLeuLysLeuLeuThrThrLeuLeuAlaValAlaAlaSer 20  
DB 1 ATGGCGGTACGCGGTTGAAGCTGTGACACACACTGCTGGCTGTCTGGCGCGTCTCC 60  
QY 21 GlnAlaGluValGluSerGluAlaGlyTyrGlyMetValThrProAspLeuPheAla 40  
DB 61 CAAGCCGAGGTGAGTCCGAGGAGGATGGGCACTGGGCGTGTCTGAGCATGGAACGG 120  
QY 41 GluGlyThrAlaAlaTyrAlaArgGlyAspTrpProGlyValValLeuSerMetGluArg 60  
DB 121 GAGGGGACCGCAGCTACGCGCGGGGACTGGCCCGGGTGGTCTCTGAGCATGGAACGG 180  
QY 61 AlaLeuArgSerArgAlaAlaLeuArgAlaLeuArgLeuArgCysArgThrGlnCysAla 80  
DB 181 GCGTGTGCGCTCCCGGAGCAGCCCTCCGCGCCCTTCCGCTGCGCTCCGACCCAGTGTGCC 240  
QY 81 AlaAppPheProTyrGluLeuAspProAspTrpSerProAlaGlnAlaSerGly 100  
DB 241 GCCGACTTCCCGTGGGAGCTGGAACCCGACTGTGTCTCCCGAGCCCGCCAGGCTCGGGC 300



101	AlaGlyAlaLeuArgAspLeuSerPhePheGlyGlyLeuLeuArgArgAlaAlaCysLeu	120	Qy
301	GC CGCGCGCCTCGCGAGCTGAGCTTCTCGGGGCGCTTCTGCGTCGGCGCTGCGCTG	360	Db
121	ArgArgCysLeuGlyProProAlaAlaHisSerLeuSerGluGluMetGluLeuGluPhe	140	Qy
361	CGCCGCTGCTCGGCGCGCGCGCCGCCACTCGCTCAGCGAAGAGATGGAGCTGGAGTTC	420	Db
141	ArgLysArgSerProTyrAsnTyrLeuGlnValAlaTyrPheLysIleAsnLysLeuGlu	160	Qy
421	CGCAAGCGGAGCCCTACAACCTACCTGCAGGTCGCTACTTCAAGATCAACAAGTTGGAG	480	Db
161	LysAlaValAlaAlaAlaAlaHisThrPhePheValGlyAsnProGluHisMetGluMetGln	180	Qy
481	AAAGCTGTTGCTGCAGACACACACCTTCTTCGTGGGCAATCCCTGAGCACATGGAATGAG	540	Db
181	GlnAsnLeuAspTyrTyrGlnThrMetSerGlyValLysGluAlaAspPheLysAspLeu	200	Qy
541	CAGAACCTAGACTATTACCAACCATGCTCGAGTGAAGGCGCGACTTCAAGATCTT	600	Db
201	GluThrGlnProHisMetGlnGluPheArgLeuGlyValArgLeuTyrSerGluGluGln	220	Qy
601	GAGACTCAACCCCATATGCAAGAATTTCGACTGGGAGTCGACTCTACTCAGAGGAACAG	660	Db
221	ProGlnGluAlaValProHisLeuGluAlaAlaLeuGlnGluTyrPheValAlaTyrGlu	240	Qy
661	CCACAGERAAGCTGTGCCCCACCTAGAGGCGCGCTGCAAGAATACTTTGTGCCCTATGAG	720	Db
241	GluCysArgAlaLeuCysGluGlyProTyrAspTyrAspGlyTyrAsnTyrLeuGluTyr	260	Qy
721	GAGTGCCTGCGCCTCGCAAGCGCCCTATGACTACGATGGCTACAACTACCTTGAGTAC	780	Db
261	AsnAlaAspLeuPheGlnAlaIleThrAspHisTyrIleGlnValLeuAsnCysLysGln	280	Qy
781	AACGCTGACCTCTTCAGGCCATCACAGATCATATCATCCAGGTCCTCAACTGTAAAGCAG	840	Db
281	AsnCysValThrGluLeuAlaSerHisProSerArgGluGlyProPheGluAspPheLeu	300	Qy
841	AACTGTGTCCAGAGCTTGCTTCCCAACCAAGTCGAGAGAAGCCCTTTGAAGACTTCTTC	900	Db
301	ProSerHisTyrAsnTyrLeuGlnPheAlaTyrTyrAsnIleGlyAsnTyrThrGlnAla	320	Qy
901	CCATCGCATATATAATTATCTGCGAGTTGCTCTACTATAAC-----	939	Db
321	GlyGluCysAlaLysThrTyrLeuLeuPhePheProAsnAspGluValMetAsnGlnAsn	340	Qy
940	-----AAGACA-----ATCTGCTATTGTAAT	960	Db
341	LeuAlaTyrTyrAlaAlaMetLeuGlyGluGluHisThrArgSerIleGlyProArgGlu	360	Qy
961	CTTCTCTGTCTCTGAAAATCTATAGAAA-----AAGAAG	996	Db
361	SerAlaLysGluTyrArgGlnArgSerLeuLeuGluLysGluLeuLeuPheAlaTyr	380	Qy
997	AGTGCCCAAGAGGATCCGACACGAGCCCTACTCGAAAAAGAACTGCTTTCTTCGCTTAT	1056	Db
381	AspValPheGlyIleProPheValAspProAspSerThrProGluGluValIlePro	400	Qy
1057	GATGTGTTTTGGAATTCCTTTGTGGATCCGGATTTCATGGACTCCAGAAGAAGTGATCCC	1116	Db
401	LysArgLeuGlnGluLysGlnLysSerGluArgGluThrAlaValArgIleSerGlnGlu	420	Qy
1117	AAGAGATTGCAGAGAAACAGAGAGTCAGAACGGGMAACAGCCGTACGCATCTCCAGGAG	1176	Db
421	IleGlyAsnLeuMetLysGluIleGluThrLeuValGluGluLysThrLysGluSerLeu	440	Qy
1177	ATTGGGAACCTTATGAAGGAAATTCGAGACCCCTGTGGAGAGAGAACCAAGGAGTCACTG	1236	Db
441	AspValSerArgLeuThrArgGluGlyGlyProLeuLeuTyrGluGlyIleSerLeuThr	460	Qy
1237	GATGTGAGCAGACATGACCCGGAAAGGTGGCCCCCTGCTGTATGAGAGCATCACTCAAC	1296	Db
461	MetAsnSerLysLeuLeuAsnGlyTyrGlnArgValValMetAspGlyValIleSerAsp	480	Qy

Db	1297	ATGAATCCAAACTCTCTGAATGGTTCCAGCGGGTGGTGATGACGGCGTAATCTCTGAC	1356
Qy	481	HisGluCysGlnGlnLeuGlnArgLeuThrAsnValAlaAlaThrSerGlyAspGlyTyr	500
Db	1357	CACAGTGTTCAGAGCTGCAGAGACTGACCAATGTGGCAGCAACCTCAGAGATGGCTAC	1416
Qy	501	ArgGlyGlnThrSerProHisThrProAsnGluLysPheTyrGlyValThrValPheLys	520
Db	1417	CGGGGTTCAGACCTCCCCACATACTCCCAATGAAGAAGTTCTATGGTGTCACTGCTTCCAA	1476
Qy	521	AlaLeuLysLeuGlyGlnGluGlyLysValProLeuGlnSerAlaHisLeuTyrTyrAsn	540
Db	1477	GCCTCAAGCTGGGGCAAGAGCAAAAGTTCTCTGCAGATGCCACCCTGTACTACAAC	1536
Qy	541	ValThrGluLysValArgArgIleMetGluSerTyrPheArgLeuAspThrProLeuTyr	560
Db	1537	GTACCGGGAAGGTGGCGGCATCATGGAGTCTACTTCCGCTGGATAGCCCCCTCTAC	1596
Qy	561	PheSerTyrSerHisLeuValCysArgThrAlaIleGluGluValGlnAlaGluArgLys	580
Db	1597	TTTTTCTACTCTCATCTGGTGTGCGCACTGCCATCGAAGAGGTCCAGGCAGAGAGGAAG	1656
Qy	581	AspAspSerHisProValHisValAspAsnCysIleLeuAsnAlaGluThrLeuValCys	600
Db	1657	GATGATAGTCATCCATGTCACGFGGCAACTGCATCTGAAATGCCAGACCCCTCGTGTGT	1716
Qy	601	ValLysGluProProAlaTyrThrPheArgAspTyrSerAlaIleLeuTyrLeuAsnGly	620
Db	1717	GTCAAGAGCCCCCAGCCTACACCTTCGCGCACTACAGCGCCATCTTTTACCTAAATGG	1776
Qy	621	AspPheAspGlyGlyAsnPheTyrPheThrGluLeuAspAlaLysThrValThrAlaGlu	640
Db	1777	GACTTCGATGGCGAAACTTTTATTTCACTGAACTGGATGCCAAGACCGGTGACGGCAGAG	1836
Qy	641	ValGlnProGlnCysGlyArgAlaValGlyPheSerSerGlyThrGluAsnProHisGly	660
Db	1837	GTGCAGGCTCAGTGTGGAGAGCGGTGGGATTTCTTCTTCAGCACTGAAACCCACATGGA	1896
Qy	661	ValLysAlaValThrArgGlyGlnArgCysAlaIleAlaLeuTrpPheThrLeuAspPro	680
Db	1897	GTGAAGGCTGTCAACAGGGCGACGCGTGTGCCATCGCCCTGTGGTTCACCTTGGACCTT	1956
Qy	681	ArgHisSerGluArgAspArgValGlnAlaAspLeuValLysMetLeuPheSerPro	700
Db	1957	CGACACAGCAGCGGGACAGGGTGCAGGCAGATGACTGTGTGAAGATGCTCTTTCAGCCCA	2016
Qy	701	GluGlnMetAspLeuSerGlnGluGlnProLeuAspAlaGlnGlnGlyProProGluPro	720
Db	2017	GAAGAGATGGACCTCTCCAGGAGCAGCCCTCGATGTCGCCAGCAGGCTCCCCCGGAACCT	2076
Qy	721	AlaGlnGlnSerLeuSerGlySerGluSerLysProLysAspGluLeu	736
Db	2077	GCACAAAGTCTCTCTCAGGAGTGAATCGAAGCCCAAGGATGAGCTA	2124
RESULT 10			
US-10-045-815-5			
; Sequence 5, Application US/10045815			
; Publication No. US20020160498A1			
; GENERAL INFORMATION:			
; APPLICANT: Wadhwa, Renu			
; APPLICANT: Sugihara, Takashi			
; APPLICANT: Ohide, Akiko			
; TITLE OF INVENTION: TUMOR SUPPRESSOR GENE			
; FILE REFERENCE: 06501-091001			
; CURRENT APPLICATION NUMBER: US/10/045,815			
; CURRENT FILING DATE: 2001-10-26			
; PRIOR APPLICATION NUMBER: PCT/JP00/02731			
; PRIOR FILING DATE: 2000-04-26			
; PRIOR APPLICATION NUMBER: JP 11/118806			
; PRIOR FILING DATE: 1999-04-26			
; NUMBER OF SEQ ID NOS: 15			
; SOFTWARE: FastSeq for Windows Version 4.0			

REF ID: A66115

```

US-10-045-815-5
; Sequence 5, Application US/10045815
; Publication No. US20020160498A1
; GENERAL INFORMATION:
; APPLICANT: Wadhwa, Renu
; APPLICANT: Sugihara, Takashi
; APPLICANT: Ohide, Akiko
; TITLE OF INVENTION: TUMOR SUPPRESSOR GENE
; FILE REFERENCE: 06501-091001
; CURRENT APPLICATION NUMBER: US/10/045,815
; CURRENT FILING DATE: 2001-10-26
; PRIOR APPLICATION NUMBER: PCT/JP00/02731
; PRIOR FILING DATE: 2000-04-26
; PRIOR APPLICATION NUMBER: JP 11/118806
; PRIOR FILING DATE: 1999-04-26
; NUMBER OF SEQ ID NOS: 15
; SOFTWARE: FastSeq for Windows Version 4.0

```

```

; SEQ ID NO 5
; LENGTH: 2416
; TYPE: DNA
; ORGANISM: Mus musculus
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (12)...(2252)
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 2376
; OTHER INFORMATION: n = A,T,C or G
US-10-045-815-5

Alignment Scores:
Pred. No.: 0 Length: 2416
Score: 3323.00 Matches: 630
Percent Similarity: 92.37% Conservative: 36
Best Local Similarity: 87.38% Mismatches: 52
Query Match: 85.87% Indels: 4
DB: 14 Gaps: 2

US-10-045-815-4 (1-736) x US-10-045-815-5 (1-2416)
QY 11 ThrLeuLeuAlaValAlaAlaAlaSer---GlnAlaGluValGluSerGluAlaGly 29
DB 40 ACGATGCTAGCGGTCGCGCGCGCGCGCGCTTACGGGTTGCGGC-GAGTCTGAGCGCGGA 98
QY 30 TrpGlyMetValThrProAspLeuLeuPheAlaGluGlyThrAlaAlaTyrAlaArgGly 49
DB 99 TGGAGCGTGGCAGCCCTGACCTGCTTACCGCAGAGGGGACCGGGGCTACTCGCGCAGG 158
QY 50 AspTrpProGlyValValLeuSerMetGluArgAlaLeuArgSerArgAlaAlaLeuArg 69
DB 159 GACTGGCCCGGGTGGTCTGAACATGAGCGGGCTCTGGCTCGCGGGCGGCGCTGCGT 218
QY 70 AlaLeuArgLeuArgCysArgThrGlnCysAlaAlaAspPheProTrpGluLeuAspPro 89
DB 219 GCCCTCGCGCTGCGCTGCGCAGCAGCTGTGCCACCGAACTGCGTGGGACCGGACCTG 278
QY 90 AspTrpSerProSerPro-----AlaGlnAlaSerGlyAlaGlyAlaLeuArgAspLeu 107
DB 279 GATCTCGTCCGACCCAGCCTGAGCAGGACCGCGCGCGCGCGCTGCGACGACCTG 338
QY 108 SerPhePheGlyGlyLeuLeuArgArgAlaAlaCysLeuArgArgCysLeuGlyProPro 127
DB 339 CGCTTCTTGGAGCGGTGCTGCGCGCTGCGCGCTGCTGCGCGCTGCGTGGCGCGCGC 398
QY 128 AlaAlaHisSerLeuSerGluGluMetGluLeuGluPheArgLysArgSerProTyrAsn 147
DB 399 TCTGCCCACTTGTAGTGAAGAACTCGAGCTGAGGTTCAACAGCGGAGCCCGCTACAAC 458
QY 148 TyrLeuGlnValAlaTyrPheLysIleAsnLysLeuGluLysAlaValAlaAlaHis 167
DB 459 TACCTGAGGTGCGCTATTCAAGATAACAAGCTGGAGAAAGCTGTGGCTGGCGCACAC 518
QY 168 ThrPhePheValGluAsnProGluHisMetGluMetGlnGlnAsnLeuAspTyrTyrGln 187
DB 519 ACCTTCTTGGGCAATCTCTGAGCACAATGAGATGCGGAGAACCTCGACTATTACCAA 578
QY 188 ThrMetSerGlyValLysGluAlaAspPheLysAspLeuGluThrGlnProHisMetGln 207
DB 579 ACCATGCTCGGGTGAAGGAGGAGCAGCTTCAGGGAATCTCGAGGCCAAGCCCATATGCA 638
QY 208 GluPheArgLeuGlyValArgLeuTyrSerGluGluGlnProGlnGluAlaValProHis 227
DB 639 GAGTTTCGGCTGGGGTACGACTCTACTACAGAGGAGAACCCAGGAAGCTGTGCCCCAC 698
QY 228 LeuGluAlaAlaLeuGlnGluTyrPheValAlaTyrGluGluCysArgAlaLeuCysGlu 247
DB 699 CTGGAGCGGCACCTGCACAGAGTACTTGTGGCCGATGAGGAGTGGCTGCGCTCTGCGAA 758
QY 248 GlyProTyrAspTyrAspGlyTyrAsnTyrLeuGluTyrAsnAlaAspLeuPheGlnAla 267
DB 1839 ACCTTCCGGGAATAACAGCGCATCTTCTTACCTCAATGGCGACTTCGATGGAGGAACTTT 1898

```

```

DB 759 GGGCCCTATGACTAGCAGCGGCTACAACTACTAGACTAGCGCTGACCTCTTCCAGGCC 818
QY 268 IleThrAspHisTyrIleGlnValLeuAsnCysLysGlnAsnCysValThrGluLeuAla 287
DB 819 ATCAGAGATCATACGTCAGGCTCCTCACTGATGAGCAACTGTGTACGGAGGTGGCT 878
QY 288 SerHisProSerArgGluLysProPheGluAspPheLeuProSerHisTyrAsnTyrLeu 307
DB 879 TCCACCCCAAGTAGGAAAGCCCTTTGAAGACTTCTCCCTTCACACTATAAATACCTA 938
QY 308 GlnPheAlaTyrTyrAsnIleGlyAsnTyrThrGlnAlaGlyGluCysAlaLysThrTyr 327
DB 939 CAGTTTGCCCTACTACCACTTGGGAACATATACAAAGCTATTGAATGTCCAAGACCTAC 998
QY 328 LeuLeuPhePheProAsnAspGluValMetAsnGlnAsnLeuAlaTyrTyrAlaAlaMet 347
DB 999 CTCCTCTTCTTCCCAATGATGAGGTGATGCACCAAGATCTGGCTATTACAGCCCATG 1058
QY 348 LeuGlyGluGluHisThrArgSerIleGlyProArgGluSerAlaLysGluTyrArgGln 367
DB 1059 CTTGGAGAAAGAGCCAGCTCCATCAGCCCCAGGAGAAATGCCGAGAAATACCGAGCT 1118
QY 368 ArgSerLeuLeuGluLysGluLeuLeuPhePheAlaTyrAspValPheGlyIleProPhe 387
DB 1119 CCAAACCTGTTGAGAAAGAACTGCTTTTCTTCGCTTATGACATTTTGAATTCCTTT 1178
QY 388 ValAspProAspSerTrpThrProGluGluValIleProLysArgLeuGlnGluLysGln 407
DB 1179 GTGGATCCGATTCATGAGCTCCAGAGAGTATCCCAAGAGATTCCAGAGAAAGACAG 1238
QY 408 LysSerGluArgGluThrAlaValArgIleSerGlnGluIleGlyAsnLeuMetLysGlu 427
DB 1239 AAGTCTGAACGGAAACAGCCGCTACGCATCTCCAGAGAGATTGGGAACCTTATGAGAA 1298
QY 428 IleGluThrLeuValGluGluLysThrLysGluSerLeuAspValSerArgLeuThrArg 447
DB 1299 ATCGAGACCTTGTGGAAAGAGAACCAAGAGTCTCTGGATGTGAGCAGACTGACCCGG 1358
QY 448 GluGlyGlyProLeuLeuTyrGluGlyIleSerLeuThrMetAsnSerLysLeuLeuAsn 467
DB 1359 GAAGTGTGCTCCTGCTGATGAGGCACTAGTCTCACCATGAACTCCAAAGTCTTGAAT 1418
QY 468 GlyTyrGlnArgValValMetAspGlyValIleSerAspHisGluCysGlnGluLeuGln 487
DB 1419 GGTCTCCAGCGGTGCTGATGGTGTGATCTCTGATGATGATGATGATGATGATGATGAT 1478
QY 488 ArgLeuThrAsnValAlaAlaThrSerGlyAspGlyTyrArgGlyGlnThrSerProHis 507
DB 1479 AGACTGACCAATGCGGCGCAGCAACTTCGGGAGATGGCTACCGAGGTGAGACTCCCCAC 1538
QY 508 ThrProAsnGluLysPheTyrGlyValThrValPheLysAlaLeuLysLeuGlyGlnGlu 527
DB 1539 ACCCCAAATGAAAGTTCTATGGTGTACTGTCTCTCAAGCTCTCAAGCTCGGCGAGAA 1598
QY 528 GlyLysValProLeuGlnSerAlaHisLeuTyrTyrAsnValThrGluLysValArgArg 547
DB 1599 GGAAGAAATTCCTCTGCGAGAGTCCCGCATGACTACCAACGTCGACAGAGAGGTGCGCGC 1658
QY 548 IleMetGluSerTyrPheArgLeuAspThrProLeuTyrPheSerTyrSerHisLeuVal 567
DB 1659 GTCATGAGTCTTACTTCCGCTCGACAGCCCTCTATTCTTCTTATCCACTTCTG 1718
QY 568 CysArgThrAlaIleGluGluValGlnAlaGluArgLysAspSerHisProValHis 587
DB 1719 TGCCGCATCGCAATAGAGAGTTCACAGGCTGAGAGGAGGACAGTAGTACCCCGCTCAC 1778
QY 588 ValAspAsnCysIleLeuAsnAlaGluThrLeuValCysValLysGluProProAlaTyr 607
DB 1779 GTGGATAACTGCATCTCTGATGCGGAGCTTCAATGTGATCAAGGAGGAGCCCGCAGCAT 1838
QY 608 ThrPheArgAspTyrSerAlaIleLeuTyrIleuAsnGlyAspPheAspGlyGlyAsnPhe 627
DB 1839 ACCTTCCGGGAATAACAGCGCATCTTCTTACCTCAATGGCGACTTCGATGGAGGAACTTT 1898

```

```
QY 628 TyrPheThrGluLeuAspAlaValThrAlaGluValGlnProGlnCysGlyArg 647
Db 1899 TACTTCCACAGAACTAGATCCCAAGACTGTGCGGCAGAGGTGCAGCCCGAGTGTGGAAGG 1958
QY 648 AlaValGlyPheSerSerGlyThrGluAsnProHisGlyValValAlaValThrArgGly 667
Db 1959 GCTGTGGGAATCTCTTCTGCACTGAGAACCCACATGGAGTGTACACAGGGGG 2018
QY 668 GlnArgCysAlaValLeuThrPheThrLeuAspProArgHisSerGluArgAspArg 687
Db 2019 CAGCGCTGGCCATCGCCCTGTGTTTCAAGCTGATCTCGGCACAGTGGAGAGACAGG 2078
QY 688 ValGlnAlaAspLeuValLysMetLeuPheSerProGluGluMetAspLeuSerGln 707
Db 2079 GTGCAGGCAGATGACTGTGTGAAGATGCTGTTCAGCCCAAGAGAGTGGACTTCCCCAG 2138
QY 708 GluGlnProLeuAspAlaGlnGlnGlyProProGluProAlaGlnGlnSerLeuSerGly 727
Db 2139 GAACGCCCTTGCCTGACCCAGCAGGCTTCGCCAGAGGCTTGGAGAGAGTTCGTGATGCT 2198
QY 728 Ser 728
Db 2199 GCT 2201

RESULT 11
US-10-045-815-7
; Sequence 7, Application US/10045815
; Publication No. US20020160498A1
; GENERAL INFORMATION:
; APPLICANT: Wadhwa, Renu
; APPLICANT: Sugihara, Takashi
; TITLE OF INVENTION: TUMOR SUPPRESSOR GENE
; FILE REFERENCE: 06501-091001
; CURRENT APPLICATION NUMBER: US/10/045, 815
; CURRENT FILING DATE: 2001-10-26
; PRIOR APPLICATION NUMBER: PCT/JP00/02731
; PRIOR FILING DATE: 2000-04-26
; PRIOR APPLICATION NUMBER: JP 11/118806
; PRIOR FILING DATE: 1999-04-26
; NUMBER OF SEQ ID NOS: 15
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 7
; LENGTH: 2322
; TYPE: DNA
; ORGANISM: Mus musculus
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (12)...(1637)
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: 2282-
; OTHER INFORMATION: n = A,T,C or G
US-10-045-815-7

Alignment Scores:
Pred. No.: 0 Length: 2322
Score: 3120.00 Matches: 602
Percent Similarity: 87.97% Conservative: 34
Best Local Similarity: 83.26% Mismatches: 50
Query Match: 80.62% Indels: 38
DB: 14 Gaps: 3

US-10-045-815-4 (1-736) x US-10-045-815-7 (1-2322)
QY 11 ThrLeuLeuAlaValValAlaAlaAspSer--GlnAlaGluValGluSerGluAlaGly 29
Db 40 ACGATGCTAGCGGTGCGCGCGCGCGCGCTTACGGGTTCGGGC-GAGTCTGAGCGGGGA 98
QY 30 TrpGlyMetValThrProAspLeuLeuPheAlaGluGlyThrAlaAlaValThrArgGly 49
Db 99 TGGACGTGGCGCGCCCTGACCTGTTTACGACAGGGGACCGCGGCTTACGCGCAGG 158
```

```
QY 50 AspTrpProGlyValValLeuSerMetGluArgAlaLeuArgSerArgAlaAlaLeuArg 69
Db 159 GACTGGCCCGGGGTGGTCTGAACATGAGGCGGGCTCTGGCTCGCGGGGGCCCTCGCT 218
QY 70 AlaLeuArgLeuArgCysArgThrGlnCysAlaAlaAspPheProTrpGluLeuAspPro 89
Db 219 GCCCTCGCGCTGCGTGGGCACACGCTGTGCCACCGAAGCTGCGTGGGCACCGACCTG 278
QY 90 AspTrpSerProSerPro-----AlaGlnAlaSerGlyAlaGlyAlaLeuArgAspLeu 107
Db 279 GATCTCGTCCGGACCCCGAGCTGAGCAGGAGCCCGGGCGCGCGCCCTGCGACGACCTG 338
QY 108 SerPhePheGlyGlyLeuLeuArgAlaAlaCysLeuArgArgCysLeuGlyProPro 127
Db 339 CGCTTCTCGAGCGGTGCTGCGCGTGGCGCTGCGCTACGCGCTGCTGCGGCGCGCC 398
QY 128 AlaAlaHisSerLeuSerGluGluMetGluLeuGluPheArgLysArgSerProTyrAsn 147
Db 399 TCTGCCCACTTGTCTGAGTGAAGAACTGGACCTGGAGTTCAACAAGCGGAGCGCTAGAAC 458
QY 148 TyrLeuGlnValAlaTyrPheLysLysLeuGluLysAlaValAlaAlaAlaHis 167
Db 459 TACCTGCGAGTGGCTATTTCAAGATAAACAAGCTGGAGAAAGCTGTGGCTGCGGCACAC 518
QY 168 ThrPhePheValGlyAsnProGluHisMetGluMetGlnGlnAsnLeuAspTyrTyrGln 187
Db 519 ACCTTCTTGTGGGCAATCTCTGAGCACATGGAGATGGCGCAGAACTCGACTATTACCAA 578
QY 188 ThrMetSerGlyValLysGluAlaAspPheLysAspLeuGluThrGlnProHisMetGln 207
Db 579 ACCATGCTGGGGTGAAGGAGCGACACTTCAGGAGATCTCGAGGCAAGCCCATATGAT 638
QY 208 GluPheArgLeuGlyValArgLeuTyrSerGluGluGlnProGlnGluAlaValProHis 227
Db 639 GAGTTTCTGGCTGGGGTACGACTCTACTCAGAGGAGAGCCACAGGAAGCTGTGCGCCAC 698
QY 228 LeuGluAlaAlaLeuGlnGluTyrPheValAlaTyrGluGluCysArgAlaLeuCysGlu 247
Db 699 CTGGAGGCGGCACCTGCAAGAGTACTTTGTGCGCGATGAGAGTGCCGCTCTGCGAA 758
QY 248 GlyProTyrAspTyrAspGlyTyrAsnTyrLeuGluTyrAsnAlaAspLeuPheGlnAla 267
Db 759 GGGCCCTATGACTACGACGGCTACAACTACTAGACTACAGCGGTGACCTCTTCCAGGCC 818
QY 268 IleThrAspHisTyrIleGlnValLeuAsnCysLysGlnAsnCysValThrGluLeuAla 287
Db 819 ATCAGAGATCATTAAGTCCAGTCCAGTCCCTCACTGTAAGCAGAACTGTGTCCGAGCTG 878
QY 288 SerHisProSerArgGluLysProPheGluAspPheLeuProSerHisTyrAsnTyrLeu 307
Db 879 TCCCAACCAAGTAGGAAAGCCCTTTGAGACTTCTCCCTTACACTATATATACCTA 938
QY 308 GlnPheAlaTyrTyrAsnIleGlyAsnTyrThrGlnAlaGlyGluCysAlaLysThrTyr 327
Db 939 CAGTTTGGCTACTACAACATTTGGGAACCTATACACAAGCTATTGAATGTGCCAAGACCTAC 998
QY 328 LeuLeuPhePheProAsnAspGluValMetAsnGlnAsnLeuAlaTyrTyrAlaAlaMet 347
Db 999 CTCTCTCTTCTTCCCAATGATGAGGTGATGCACCAAGATCTGGCTTATATACACAGCCATG 1058
QY 348 LeuGlyGluGluHisThrArgSerIleGlyProArgGluSerAlaLysGluTyrArgGln 367
Db 1059 CTTGGAGAGAGAGAGCGCCAGCTCCATCAGCCCGAGGAGATGCCGAGGAATACCGAGCT 1118
QY 368 ArgSerLeuLeuGluLysGluLeuPhePheAlaTyrAspValPheGlyIleProPhe 387
Db 1119 CCAACACTGTTGGAGAAAGAACTGCTTTTCTTCGCTTATGACATTTTGGATTCCTCTTT 1178
QY 388 ValAspProAspSerTrpThrProGluGluValIleProLysArgLeuGlnGluLysGln 407
Db 1179 GTGGATCCCGATTCTGACTTCCAGAGAGTGTATCCAGAGAGATTCCAGAGAGAACAG 1238
```



572 CAGAACTAGACTATTACCAACCATGCTGGAGTGAAGAGGCCGACTTCAAGGATCTT 631  
QY GluThrGlnProHisMetGlnGluPheArgLeuGlyValArgLeuTyrSerGluGluGln 220  
Db GAGACTCAACCCCATATGCAAGAAATTCGACTGGAGTGCAGCTCTACTCAGAGGAACAG 691  
QY ProGlnGluAlaValProHisLeuGluAlaLeuGlnGluTyrPheValAlaTyrGlu 240  
Db CCACAGGAAGCTGTGCCACCTAGAGGCGCGCTCAAGAAATCTTGTGGCCTATGAG 751  
QY GluCysArgAlaLeuGluGlyProTyrAspTyrAspGlyTyrAsnTyrLeuGluTyr 260  
Db GAGTGGCTGCTCTGCGAAGGCGCTATGACTACGATGCTACAACTACCTTTGAGTAC 811  
QY AsnAlaAspLeuPheGlnAlaLeuThrAspHisTyrIleGlnValLeuAsnGlyGln 280  
Db AACGCTGACCTCTCCAGGCGCATCAGAGATCAATACATCCAGGTCTCACTGTAAGCAG 871  
QY AsnCysValThrGluLeuAlaSerHisProSerArgGluLysProPheGluAspPheLeu 300  
Db AACGTGTCTCAGGAGCTTGTCTCCACCAGTCGAGAGAGCCCTTTGAGACTTCTCTC 931  
QY ProSerHisTyrAsnTyrLeuGlnPheAlaTyrTyrAsnIleGlyAsnTyrThrGlnAla 320  
Db CCATCGCATTAATATCTGCGATTTGCTCTCTCTCTCCCAATGACGAGGTGATGAACCAAAAT 1051  
QY GlyGluCysAlaLysThrTyrLeuLeuPhePheProAsnAspGluValMetAsnGln 340  
Db GTTGAATGTGCCAGACCTATCTCTCTCTCTCCCAATGACGAGGTGATGAACCAAAAT 1051  
QY LeuAlaTyrTyrAlaAlaMetLeuGluGluHisThrArgSerIleGlyProArgGlu 360  
Db TTGGCTCTATTATGAGTATGCTTGGAGAGACACACACAGATCCATCGCCCGCTGAG 1111  
QY SerAlaLysGluTyrArgGlnArgSerLeuLeuGluLysGluLeuLeuPheAlaTyr 380  
Db AGTCCCAAGGAGTACCCAGCAGCAAGCTACTGGAAGAAAGAACTGCTTTCTCGCTTAT 1171  
QY AspValPheGlyIleProPheValAspProAspSerTrpThrProGluGluValIlePro 400  
Db GATGTTTTTGAATTCCTTTTGGATTCGGATTCAGGATTCAGGATTCAGGATTCAGGATTC 1231  
QY LysArgLeuGlnGluLysGlnLys 408  
Db AAGAGATTGCAAGAGAAACAGAG 1255

## RESULT 13

US-10-094-749-921  
; Sequence 921, Application US/10094749  
; Publication No. US20030219741A1  
; GENERAL INFORMATION:  
; APPLICANT: ISOGAI, TAKAO  
; APPLICANT: OTSUKI, TETSUJI  
; APPLICANT: WAKAMATSU, AI  
; APPLICANT: SATO, HIROYUKI  
; APPLICANT: ISHII, SHIZUKO  
; APPLICANT: YAMAMOTO, JUN-ICHI  
; APPLICANT: ISONO, YUUKO  
; APPLICANT: HIO, YURI  
; APPLICANT: OTSUKA, KAORU  
; APPLICANT: NAGAI, KEIICHI  
; APPLICANT: IRIE, RYOTARO  
; APPLICANT: TAMECHIKA, ICHIRO  
; APPLICANT: SEKI, NAOHICO  
; APPLICANT: YOSHIKAWA, TSUTOMU  
; APPLICANT: OTSUKA, MOTOKUJI  
; APPLICANT: NAGAHARI, KENJI  
; APPLICANT: MASUHO, YASUHIKO  
; TITLE OF INVENTION: NOVEL FULL-LENGTH cDNA  
; FILE REFERENCE: 084335/0160  
; CURRENT APPLICATION NUMBER: US/10/094,749  
; CURRENT FILING DATE: 2002-03-12

; PRIOR APPLICATION NUMBER: 60/350,435  
; PRIOR FILING DATE: 2002-01-24  
; PRIOR APPLICATION NUMBER: JP 2001-328391  
; PRIOR FILING DATE: 2001-09-14  
; NUMBER OF SEQ ID NOS: 3381  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 921  
; LENGTH: 3396  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
; US-10-094-749-921  
  
Alignment Scores:  
Pred. No.: 2,88e-184 Length: 3396  
Score: 1751.00 Matches: 356  
Percent Similarity: 63.52% Conservative: 102  
Best Local Similarity: 49.38% Mismatches: 211  
Query Match: 45.25% Indels: 52  
DB: 16 Gaps: 12  
  
US-10-045-815-4 (1-736) x US-10-094-749-921 (1-3396)  
QY 23 GluValGluSerGluAlaGlyTrpGlyMetValThrPro---AspLeuLeuPheAlaGlu 41  
Db 295 GAGCTGGAGCTGGAGCCC-----GGGCTCTCTGAGCCCTTCGACCTGCTCTACCCGACG 348  
QY 42 GlyThrAlaAlaTyrAlaArgGlyAspTrpProGlyValValLeuSerMetGluAla 61  
Db 349 GCGCGGCGCGCTACTACAGCGGAGACTACGAGCGCGGTGGCGACTTGGAGCGGCG 408  
QY 62 LeuArgSerArgAlaAlaLeuArgAlaLeuArgLeuArgCysArgThrGlnCysAlaAla 81  
Db 409 CTCGCGAGCCACCGCGCGCTGCGGAAATTCGACGCGCTGTGCCCGCCACTGCGCGGCG 468  
QY 82 AspPheProTrpGluLeuAspProAspTrpSerProAlaGlnAlaSerGlyAla 101  
Db 469 CGCCACCCG-----CTCCGCGCCCGCCCGCCCGCGGCGGCCCC 507  
QY 102 GlyAlaLeuArgAspLeuSerPheGlyGlyLeuLeuArgArgAlaAlaCysLeuArg 121  
Db 508 GCGGCT-----GAGCTGCCCCCTTTTCGCTCTCTTGGGGCGGCGCGCTGTTATCGC 561  
QY 122 ArgCys-----LeuGlyProProAlaAla---HisSerLeuSerGluGluMet 136  
Db 562 AGCTGTGAGACCCAGCGCTCGGGGCGCCCGCATCCGCCACCGGTACAGCGGATGTG 621  
QY 137 GluLeuGluPheArgLysArgSerProTyrAsnTyrLeuGlnValAlaTyrPheLysIle 156  
Db 622 CGCAGCGACTTCCAGCGCAGAGTGCCCTACAACTACCTGCGCGGCGCTTACATCAAGCTT 681  
QY 157 AsnLysLeuGluLysAlaValAlaAlaAlaHisThrPhePheValGlyAsnProGluHis 176  
Db 682 AACCACTCGAAAGACAGTGAAGCAGCTACACATTTTCGTGGCTAACCTGAGCAC 741  
QY 177 MetGluMetGlnGlnAsnLeuAspTyrTyrGlnThrMetSerGlyValLysGluAlaAsp 196  
Db 742 ATGGAATGCGACGAGAACATTGAGATTACAGGCGGCGACAGCTGGTGTGAAGCATTCAG 801  
QY 197 PheLysAspLeuGluThrGlnProHisMetGlnGluPheArgLeuGlyValArgLeuTyr 216  
Db 802 TTGCTAGACAGAGAGCCAGCCACATGAGAGTTCACATGCGAGGAGTTAAACATTTAT 861  
QY 217 SerGluGluGlnProGlnGluAlaValProHisLeuGluAlaAlaLeuGlnGluTyrPhe 236  
Db 862 GAGGCTGAGCTTTGAGATGGCTATCAGGCATTTGGAACAAGCTTTAAGAGATATTTC 921  
QY 237 ValAlaTyrGluGluCysArgAlaLeuGluGlyProTyrAspTyrAspGlyTyrAsn 256  
Db 922 GTTGAAGATACAGAAATGCCGCGACCTATGAGGGGCGCTCAGAGATTTGAAGAAATATGAG 981  
QY 257 TyrLeuGluTyrAsnAlaAspLeuPheGlnAlaIleThrAspHisTyrIleGlnValLeu 276  
Db 982 TATTTAGGGTATAAGGCTGCTGTGATGAAGCTATTTCGAGATCCTACATCAGGAGTCTT 1041

QY 277 AsnCysLysGlnAsnCysValThrGluLeuAlaSerHisProSerArgGluLysProPhe 296  
 Db 1042 GTTGTGACAGTAATGTTGAGGAGACTTGCACCCGCCCTTCCGCCCTCTCTCCCATC 1101  
 QY 297 GluAspPheLeuProSerHisTyrAsnTyrLeuGlnPheAlaTyrTyrAsnIleGlyAsn 316  
 Db 1102 GAGAAATTTCTCTCTCTGCACTATGATTACCTACACAGTTTGCCTACTATCGAGTTGGTGAG 1161  
 QY 317 TyrThrGlnAlaGlyGluCysAlaLysThrTyrLeuLeuPhePheProAsnAspGluVal 336  
 Db 1162 TATGTGAAGCCCTCGAGTGTGCCAAGCCCTATCTCTATGCCATCCAGATGATGAGGAT 1221  
 QY 337 MetAsnGlnAsnLeuAlaTyrTyrAlaAlaMetLeuGlyGluGlu---HisThrArgSer 355  
 Db 1222 GTCCCTAGACAATGTGATTACTATGAGAGTCTGCTGGATGATAGCATTTGACCCGCCATCC 1281  
 QY 356 IleGlyProArgGluSerAlaLysGluTyrArgGlnArgSerLeuLeuGluLysGluLeu 375  
 Db 1282 ATTGAGGCCAGAGAGGATTTAACATGTTTGTGAACGTCATAAGCTGAGCTCTGAGCTG 1341  
 QY 376 LeuPhePheAlaTyrAspValPheGlyIleProPheValAspProAspSerTyrThrPro 395  
 Db 1342 ATAAATCAGCTGCAGAGGCTGCGGTTTTCATACACTGAACCAATTTATGATC--- 1398  
 QY 396 GluGluValIleProLysArgLeuGlnGluLysGlnLysSerGluArgGluThrAlaVal 415  
 Db 1399 -----AGATATGAGGAGCAGCAGGATGAGAAT 1425  
 QY 416 ArgIleSerGlnGluIleGlyAsnLeuMetLysGluIleGluThrLeuValGluGluLys 435  
 Db 1426 CGGGTCCCTCAGGAGTGAACGTAGAGGAGCAGAGTTCATGATTCTCAATGGGAAAA 1485  
 QY 436 ThrLysGluSerLeuAspValSerArgLeuThrArgGluGlyGlyProLeuLeuTyrGlu 455  
 Db 1486 ---AAGCTATCACCAAGATAGTCGACCTAAGAGAGGTGCTCTCTACTCTATGAG 1542  
 QY 456 GlyIleSerLeuThrMetAsnSerLysLeuLeuAsnGlyTyrGlnArgValValMetAsp 475  
 Db 1543 AACATCATCATTCGTCTCAACTCGAGCAGCTGAACGGGACTCAGCGGGTCTCTCTGGAT 1602  
 QY 476 GlyValIleSerAspHisGluCysGlnGluLeuGlnArgLeuThrAsnValAlaAlaThr 495  
 Db 1603 AACGCTCTGTGGAGAACAGCTGCCAGAGCTCCACAGCTGGCCAGTGAATCATGCTT 1662  
 QY 496 SerGlyAspGlyTyrArgGlyGlnThrSerProHisThrProAsnGluLysPheTyrGly 515  
 Db 1663 GTTGTGATGGATACAGAGGAAAACTTCACCCCATACACCAATGAAAGTTTGAAGGT 1722  
 QY 516 ValThrValPheLysAlaLeuLysLysLeuGlnGluLysValProLeuGlnSerAla 535  
 Db 1723 GCAACTGTCTGAAAGCACTCAATCTGTTATGAAGTCCAGTCCACCTGAAGAGCGCT 1782  
 QY 536 HisLeuTyrTyrAsnValThrGluLysValArgIleMetGluSerTyrPheArgLeu 555  
 Db 1783 CGTCTGTTTATGATACAGCGAAAGCGCTCGAAGGATTGAGATCTTATTTATGCTG 1842  
 QY 556 AspThrProLeuTyrPheSerTyrSerHisLeuValCysArgThrAlaIleGluGluVal 575  
 Db 1843 AACTCAACTCTGTATTTTCTATACACATGCTTCGCGAAGCAGCCCTGTCTGTCTAG 1902  
 QY 576 GlnAlaGluArgLysAspSerHisProValHisValAspAsnCysIleLeuAsnAla 595  
 Db 1903 CAGGATAGAAGAAATGACCTCAGTCATCCCATCCATGCTGACAACTGTTTGTGGATCCA 1962  
 QY 596 GluThrLeuValCysValLysGluProProAlaTyrThrPheArgAspTyrSerAlaIle 615  
 Db 1963 GAGGCCAACGAATGCTCGAAGAGGAGCCTCTGCTTACACATTTTCAGACTATAGTCTCTC 2022  
 QY 616 LeuTyrLeuAsnGlyAspPheAspGlyGlyAsnPheTyrPheThrGluLeuAspAlaLys 635  
 Db 2023 CTATATATGATGATGATCTTTGAAGGAGGAGATTTCTATTTTCACAGAGATGATGCTAG 2082

QY 636 ThrValThrAlaGluValGlnProGlnCysGlyArgAlaValAlaGlyPheSerSerGlyThr 655  
 Db 2083 ACTGTGACTGCTCTATATAAACCAAAATGTGGCGCATGATCAGCTTCTCATCTCGAGA 2142  
 QY 656 GluAsnProHisGlyValLysAlaValThrArgGlyGlnArgCysAlaIleAlaLeuTyr 675  
 Db 2143 GAGAACCCTCATGGGTGAAGCAGCTCACCAAGGGAAGAGGTGTGCTGTGCTCTGTGG 2202  
 QY 676 PheThrLeuAspProArgHisSerGluArgAspArgValGlnAlaAspLeuValLys 695  
 Db 2203 TTCACCTTGGACCCACTTTATAGAGAAATTGGAGCAATACAGGCTGATGAAGTATGCA 2262  
 QY 696 MetLeuPheSerProGluGluMetAspLeuSerGlnGluGlnProLeuAspAlaGln 715  
 Db 2263 ATT-----CTGGATCAAGAA-----CAGCA 2283  
 QY 716 GlyProProGluProAlaGlnGluSerLeuSerGlySerGluSerLysProLysAspGlu 735  
 Db 2284 GCGAAGCATGAA-----CTGAATATCAACCCCTAAAGATGAG 2319  
 QY 736 Leu 736  
 Db 2320 CTA 2322

RESULT 14  
 US-10-071-766-86  
 ; Sequence 86, Application US/10071766  
 ; Publication No. US20020192678A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Huel-Wei Chen  
 ; TITLE OF INVENTION: GENES EXPRESSED IN SENESENCE  
 ; FILE REFERENCE: PA-0043 US  
 ; CURRENT APPLICATION NUMBER: US/10/071,766  
 ; CURRENT FILING DATE: 2002-02-07  
 ; NUMBER OF SEQ ID NOS: 144  
 ; SOFTWARE: PERL Program  
 ; SEQ ID NO 86  
 ; LENGTH: 2665  
 ; TYPE: DNA  
 ; ORGANISM: Homo sapiens  
 ; FEATURE:  
 ; NAME/KEY: misc feature  
 ; OTHER INFORMATION: Inocyte ID No. US20020192678A1 978730.4  
 ; NAME/KEY: unsure  
 ; LOCATION: 353-387, 694, 727  
 ; OTHER INFORMATION: a, t, c, g, or other  
 US-10-071-766-86

Alignment Scores:  
 Pred. No.: 1,8e-180 Length: 2665  
 Score: 1715.50 Matches: 356  
 Percent Similarity: 63.49% Conservative: 103  
 Best Local Similarity: 49.24% Mismatches: 211  
 Query Match: 44.33% Indels: 53  
 DB: 14 Gaps: 12

US-10-045-815-4 (1-736) x US-10-071-766-86 (1-2665)  
 QY 23 GluValGluSerGluAlaGlyTyrGlyMetValThrPro---AspLeuLeuPheAlaGlu 41  
 Db 430 GAGCTGGAGCTGGAGCCC-----GGGCCTCTGCGACCTTCGACCTCTCTACGCCAGC 483  
 QY 42 -GlyThrAlaAlaTyrAlaArgGlyAspTyrProGlyValValLeuSerMetGluArgAl 61  
 Db 484 GGGCGCGGCCCTCTACTACCGGAGACTACGACGAGCGGTGCGCGACTTGGAGCGGC 543  
 QY 61 aLeuArgSerArgAlaAlaLeuArgAlaLeuArgLeuArgCysArg-ThrGlnCysAlaA 81  
 Db 544 GCTCGCAGCAGCACCGCGCCCTCGCGAAATCCGACACGCGCTGTCCCTGCACCTGCCGCG 603  
 QY 81 laAspPheProTyrGluLeuAspProAspTyrPheProSerProAlaGlnAlaSerGlyA 101  
 Db 604 CGCGCCACCCG-----TCTCCGCCCGCCCGCCCGCGCGAGGGG 642



Qy	101	laGlyAlaLeuArgAspLeuSerPhePheGlyGlyLeuLeuArgArgAlaAlaCysLeuA	121
Db	643	CCCCGGCGCTG--AGCTGCCCCCTTTTCGCTCCTTGTGGGGCGGCGCGCTTTATC	699
Qy	121	rgArgCys-----LeuGlyProProAlaAla---HisSerLeuSerGluGlu	136
Db	700	GCAGCTGTGAGACCCAGCGCTCGGGGNCCTCCGATCCCGCCACCCGCGCAGCGAGATG	759
Qy	136	etGluLeuGluPheArgLysArgSerProTyrAsnTyrLeuGlnValAlaTyrPheLysI	156
Db	760	TGGCAGCGACTTCCAGCGCAGAGTGCCCTTACAACTACCTGACGGCGGCTCATCAAGC	819
Qy	156	leAsnLysLeuGluLysAlaValaAlaAlaAlaHisThrPhePheValGlyAsnProGluH	176
Db	820	TTAACCAAGCTCGAAAAGCAGTGAGGAGCTCACACATTTTCGTGGCTTAACCTGAGC	879
Qy	176	isMetGluMetGlnGlnAsnLeuAspTyrTyrGlnThrMetSerGlyValLysGluAlaA	196
Db	880	ACATGGAAATGCGACAGAACTAGAGAAATACAGGGCGACAGCTGGTGTGGAAGCATTCG	939
Qy	196	spPheLysAspLeuGluThrGlnProHisMetGlnGluPheArgLeuGlyValArgLeuT	216
Db	940	AGTTGGTAGACAGAGAGCCAGCCACACATGGAGAGTTACAATGCAGGAGTTAAACATT	999
Qy	216	YrSerGluGluGlnProGlnGluAlaValProHisLeuGluAlaAlaLeuGlnGluTyrP	236
Db	1000	ATGAGGCTGATGACTTTCAGATGGCTATCAGGCACCTTCGAACAAGCGCTTAAGAGAATATT	1059
Qy	236	heValAlaTyrGluGluCysArgAlaLeuCysGluGlyProTyrAspTyrAspGlyTyrA	256
Db	1060	TCGTTGAAGATACAGAATGCCGACCTATGTGAGGGGCTCAGAGATTGGAAGAAATATG	1119
Qy	256	snTyrLeuGluTyrAsnAlaAspLeuPheGlnAlaIleThrAspHisTyrIleGlnVal	276
Db	1120	AGTATTTAGGCTAATAGGCTGGTCTGTATGAAGCTATTGCAGATCACTACATCAGGTGC	1179
Qy	276	euAsnCysLysGlnAsnCysValThrGluLeuAlaSerHisProSerArgGluLysProP	296
Db	1180	TTGTTTGTGACATGAATGTGTGAGGAACTTGCACCCCGCTGGCGGCTCTCTCCCA	1239
Qy	296	heGluAspPheLeuProSerHisTyrAsnTyrLeuGlnPheAlaTyrTyrAsnIleGlyA	316
Db	1240	TCGAGAATTTTCTTCCTCTGCACATGATACCTACAGTTTGCTACTACTCGAGTTGGTG	1299
Qy	316	snTyrThrGlnAlaGlyCysAlaLysThrTyrLeuLeuPhePheProAsnAspGluV	336
Db	1300	AGTATGTGAAGCCCTGGAGTGTGCCAAGGCTATCTTCTATGCCATCCAGATGATGAGG	1359
Qy	336	alMetAsnGlnAsnLeuAlaTyrTyrAlaAlaMetLeuGlyGluGlu---HisThrArgS	355
Db	1360	ATGTCTCTAGACAATGTGGATTACTATGAGCTCTGCTGGATGATGAGCATTTGACCCGCGAT	1419
Qy	355	erIleGlyProArgGluSerAlaLysGluTyrArgGlnArgSerLeuLeuGluLysGluL	375
Db	1420	CCATTGAGCCAGAGAGATTAAACATGTTTGTGAAACGCTCATAAAGCTGGAGCTGAGC	1479
Qy	375	euLeuPhePheAlaTyrAspValPheGlyIleProPheValAspProAspSerTyrThrP	395
Db	1480	TCATAAAATCAGCTCGACAGAGGCTCTGGGGTTTTCATACACTGAACCGAAATATTGGATC-	1538
Qy	395	roGluGluValIleProLysArgLeuGlnLysGlnLysSerGluArgLeuThrAlaV	415
Db	1539	-----AGATATGGAGGACGACAGATGAGA	1563
Qy	415	alArgIleSerGlnGluIleGlyAsnLeuMetLysGluIleGluThrLeuValGluGluL	435
Db	1564	ATCGGGTCCCTTCAGGAGTGAACGATAGAGGAGCAGAGTAAGTTTCATGGATTCTCAATGGAA	1623
Qy	435	yeThrLysGluSerLeuAspValSerArgLeuThrArgGluGlyGlyProLeuLeuTyrG	455
Db	1624	AA---AAGCTATACCCCAAGATAGATCGACAGCTTAAGAGAGGTGGTCTCTACTCTATG	1680





Db	1161	ATCCAGCGCTTCATCTCCGATCCCTGGGGGAGAGAGGAGCTCTACTATGCCATGGAG	1220
Qy	382	ValPheGlyLeuProPheValAspProAspSerThrProGluGluValIleProLys	401
Db	1221	CACCTGGGGACCAAGCTCAAGATCTGACCCCTGGACCCCTGCAGCTCTCATCCTCGAG	1280
Qy	402	ArgLeuGluGluLysGlnLysSerGluArgGluThrAlaValArgIleSerGlnGluIle	421
Db	1281	GCACCTTAGAGAAAGCTCAGAGAGGATCAGAGAAG-----	1316
Qy	422	GlyAsnLeuMetLysGluIleGluThrLeuValGluGluLysThrLysGluSerLeuAsp	441
Db	1317	-----AGGCCTTGGGACCATGAGCCCGTGAAG	1343
Qy	442	ValSerArgLeuThrArgGluGlyGlyProLeuLeuTyGluGlyIleSerLeuThrMet	461
Db	1344	CCAAAGCCCTTGACCTACTGGAAGGATGCTCTCTCTGGAGGGTGTGACCTTGACCCAG	1403
Qy	462	AsnSerLysLeuLeuAsnGlyTyGlnArgValValMetAspGlyValIleSerAspHis	481
Db	1404	GATTCAGGAGCTGATGGTGGAGCGGCGGTGTGGATGGGTGCTCACCCCCAGCC	1463
Qy	482	GluCysGlnGluLeuGlnArgLeuThrAsnValAlaAlaThrSerGly-----	499
Db	1464	GAGTGTGGGTGTGCTGTCAGCTGGCTAAGGATGCAGCTGGGCTGGAGCCAGGTCTGGC	1523
Qy	500	TyrArgGlyGlnThrSerProHisThrProAsnGluLysPheTyrGlyValThrValPhe	519
Db	1524	TATCGTGTGCGGCTCCCTCACACCCCATGAGCGTTCGAGGGGCTCACGGTGCTT	1583
Qy	520	LysAlaLeuLysLeuGlyGlnGluGlyLysValProLeuGlnSerAlaHisLeuTyrTyr	539
Db	1584	AAGGCTCGCAGCTGGCCCGGCTGGGACAGTGGGCGAGTCAGGCTGCTAAGCTGCTTCTG	1643
Qy	540	AsnValThrGluLysValArgAlaGlyIleMetGluSerTyrPheArgLeuAspThrProLeu	559
Db	1644	GAGGTGAGCGGGGTGGCGACCTTGACCCAGCCCTACTTCCCGGAAACGGCCCTCG	1703
Qy	560	TyrPheSerTyrSerHisLeuValCysArgThrAlaIleGluGluValGlnAlaGluArg	579
Db	1704	CATCTGTCCTTACCCACCTGGTGTGCGCGAGGCCATAGAGGAGAGCAAGAGCAGCGC	1763
Qy	580	LysAspSerHisProValHisValAspAsnCysIleLeuAsnAlaGluThrLeuVal	599
Db	1764	ATGACCTGATGATCCACAGTGCACGCAGACAACCTGCGTCTGGACCTCCTGACCGGAGAG	1823
Qy	600	CysValLysGluProProAlaTyrThrPheArgAspTyrSerAlaIleLeuTyrLeuAsn	619
Db	1824	TGCTGGGGGAGCCCCAGCCTACACCTATCGGACTACAGCGGACTCTCTTACCTCAAC	1883
Qy	620	GlyAspPheAspGlyGlyAsnPheTyrPheThrGluLeuAspAlaLysThrValThrAla	639
Db	1884	GATGACTTCCAGGTGGGACCTGTCTTACGGAGGCCAACGCCCTCACTGTCACGGCT	1943
Qy	640	GluValGlnProGlnCysGlyArgAlaValGlyPheSerSerGlyThrGluAsnProHis	659
Db	1944	CGGGTGGCGTCTCGCTGGGGCGCTGTGGCCCTTACGCTCCGGTGTGAGATCCCAT	2003
Qy	660	GlyValLysAlaValThrArgGlyGlnArgCysAlaIleAlaLeuTrpPheThrLeuAsp	679
Db	2004	GGGTGTGGCGGTGACTCGGGGACGGCGCTGTGCCCTGGCACTGTGGCACACGTGGGCA	2063
Qy	680	ProArgHisSerGluArgAspArgValGlnAlaAspAspLeuValLysMetLeuPheSer	699
Db	2064	CCTGAGCACAGGAGCAGGAGGTGATGAGAAAGAAATGCCACAAAGACCTTCCCCAGAG	2114
Qy	700	ProGluGluMetAspLeuSerGlnGluGlnProLeuAspAlaGlnGlnGlyProProGlu	719
Db	2115	TCACAGGAGGAGGAGAGAGAGAGAAATGCCACAAAGACCTTCCCCAGAG	2174
Qy	720	ProAlaGln-----GluSerLeuSerGlySerGluSerLysProLys	733
Db	2175	CCCCCTACCGCAGGCACAGAGGCTCCAGACAGACTGGAGAGGCGACCTCGGGTTCCG	2234

Qy 734 AspGluLeu 736  
Db 2235 GAGGAGCTG 2243

Search completed: July 18, 2004, 18:22:27  
Job time : 1163 secs

**This Page Blank (uspto)**

OM protein - nucleic search, using frame\_plus\_p2n model

Run on: July 18, 2004, 09:59:50 ; Search time 20888 Seconds

(without alignments)  
1527.215 Million cell updates/sec

US-10-045-815-4

Title:

Perfect score:

Sequence:

1 MAVRALKLTLLAVVAAS.....PPEPAQESLSGSESKPKDEL 736

Scoring table:

BLOSUM62  
Xgapop 10.0 , Xgapext 0.5  
Ygapop 10.0 , Ygapext 0.5  
Fgapop 6.0 , Fgapext 7.0  
Delop 6.0 , Delext 7.0

Searched: 3470272 seqs, 21671516995 residues

Total number of hits satisfying chosen parameters: 6940544

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:

-MODEL=frame+ p2n.model -DEV=xlp  
-Q=/cgn2\_1/USPTO\_specol\_p/US10045815/runat\_14072004\_123008\_16818/app\_query.fasta\_1.903  
-DB=genembl -QPMF=fastap -SUFFIX=std.rge -MINMATCH=0.1 -LCORCL=0 -LCORPEXT=0  
-UNITS=bits -START=1 -END=1 -MATRIX=BLOSUM62 -TRANS=human40.cdd -LIST=45  
-OUTFMT=200 -THR\_SCORE=pct -THR\_MAX=100 -THR\_MIN=0 -ALIGN=15 -MODE=LOCAL  
-DOCALLIGN=200 -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000  
-USER=US10045815 @CGN 1 1 7406 @runat\_14072004\_123008\_16818 -NCPU=6 -ICPU=3  
-NO MMAP -LARGEQUERY -NEG\_SCORES=0 -WAIT -DGPBLOCK=100 -LONGLOG  
-DEV\_TIMEOUT=120 -WARN\_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6  
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :

GenEmbl:

1: gb\_ba:\*  
2: gb\_htg:\*  
3: gb\_in:\*  
4: gb\_em:\*  
5: gb\_ov:\*  
6: gb\_pat:\*  
7: gb\_ph:\*  
8: gb\_pl:\*  
9: gb\_pr:\*  
10: gb\_ro:\*  
11: gb\_sts:\*  
12: gb\_sy:\*  
13: gb\_un:\*  
14: gb\_vi:\*  
15: em\_ba:\*  
16: em\_fun:\*  
17: em\_hum:\*  
18: em\_in:\*  
19: em\_mu:\*  
20: em\_om:\*  
21: em\_or:\*  
22: em\_ov:\*  
23: em\_pat:\*  
24: em\_ph:\*  
25: em\_pl:\*  
26: em\_ro:\*  
27: em\_sts:\*  
28: em\_un:\*

29: em\_vi:\*  
30: em\_htg\_hum:\*  
31: em\_htg\_inv:\*  
32: em\_htg\_other:\*  
33: em\_htg\_mus:\*  
34: em\_htg\_pln:\*  
35: em\_htg\_rod:\*  
36: em\_htg\_mam:\*  
37: em\_htg\_vrt:\*  
38: em\_sy:\*  
39: em\_htgo\_hum:\*  
40: em\_htgo\_mus:\*  
41: em\_htgo\_other:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match %	Query Length	DB ID	Description
1	3870	100.0	2600	9 AF097432	AF097432 Homo sapi
2	3846	99.4	2563	6 AX136191	AX136191 Sequence
3	3846	99.4	2563	6 BD123548	BD123548 Secretary
4	3846	99.4	2563	9 AK075418	AK075418 Homo sapi
5	3846	99.4	2583	6 AX704765	AX704765 Sequence
6	3846	99.4	2585	6 AX877288	AX877288 Sequence
7	3846	99.4	2585	6 BD156581	BD156581 Primer fo
8	3846	99.4	2585	9 AK027697	AK027697 Homo sapi
9	3844	99.3	2524	6 AX877139	AX877139 Sequence
10	3844	99.3	2524	6 BD156505	BD156505 Primer fo
11	3844	99.3	2524	9 AK027680	AK027680 Homo sapi
12	3780	97.7	2829	9 AF097431	AF097431 Homo sapi
13	3726.5	96.3	2993	6 AX880254	AX880254 Sequence
14	3726.5	96.3	2993	6 BD158280	BD158280 Primer fo
15	3726.5	96.3	2993	9 AK027648	AK027648 Homo sapi
16	3403.5	87.9	2869	10 BC024047	BC024047 Mus muscu
17	3383.5	87.4	2524	10 AF087433	AF087433 Rattus no
18	3323	85.9	2391	10 AF165163	AF165163 Mus muscu
19	3120	80.6	2297	10 AF165164	AF165164 Mus muscu
20	2085	53.9	1562	9 BC015309	BC015309 Homo sapi
21	2024	52.3	1173	9 BT007039	BT007039 Homo sapi
22	2024	52.3	1173	12 BT007768	BT007768 Synthetic
23	2022	52.2	1526	9 AK025841	AK025841 Homo sapi
24	1751	45.2	3396	6 AX714237	AX714237 Sequence
25	1751	45.2	3396	9 AK056447	AK056447 Homo sapi
26	1747	45.1	2127	6 AX815645	AX815645 Sequence
27	1747	45.1	2261	6 AX815648	AX815648 Sequence
28	1747	45.1	3396	9 HS2430351	AJ430351 Homo sapi
29	1747	45.1	3496	9 AK125334	AK125334 Homo sapi
30	1715.5	44.3	2249	10 MMU430350	AJ430350 Mus muscu
31	1575.5	40.7	2789	6 AX877037	AX877037 Sequence
32	1575.5	40.7	2789	6 BD156451	BD156451 Primer fo
33	1575.5	40.7	2789	9 AK001580	AK001580 Homo sapi
34	1510	39.0	2801	9 BC005029	BC005029 Homo sapi
35	1473	38.1	2848	9 AK025976	AK025976 Homo sapi
36	1473	38.1	2793	9 HSA430349	AJ430349 Homo sapi
37	1422	36.7	2779	10 MMU441086	AJ441086 Mus muscu
38	1318.5	34.1	2261	9 BC017217	BC017217 Homo sapi
39	1271	32.8	2289	10 BC003726	BC003726 Mus muscu
40	1247.5	32.2	2129	6 AX772981	AX772981 Sequence
41	1247.5	32.2	2129	9 HSU47926	U47926 Human unkno
42	1190	30.7	2214	10 BC016431	BC016431 Mus muscu
43	1179.5	30.5	1656	6 AX772984	AX772984 Sequence
44	1169	30.2	3226	9 AK126766	AK126766 Homo sapi
45	1108	28.6	703	6 AX868722	AX868722 Sequence

ALIGNMENTS

RESULT 1



Db 1312 ATTGGGAACCTTATGAGGAATCGAGACCCCTTGTGGAAGAGACCAAGAGTCACTG 1371  
Qy 441 AspValSerArgLeuThrArgGluGlyGlyProLeuLeuTyrGluGlyLeuSerLeuThr 460  
Db 1372 GATGTGACGACATGACCCGGGAAGTGGCCCTGCTGTATGAAGGCATCACTCAAC 1431  
Qy 461 MetAsnSerLysLeuLeuAsnGlyTyrGlnArgValMetAspGlyValIleSerAsp 480  
Db 1432 ATGACTCCAACTCTCTGAATGTTACACGCGGTGTGTATGACGGGTAACTCTGAC 1491  
Qy 481 HisGluCysGlnGluLeuGlnArgLeuThrAsnValAlaAlaThrSerGlyAspGlyTyr 500  
Db 1492 CACGAGTGTGAGGAGTGCAGAGACTGACCAATGTGCGCAACCTCAGGAGATGCTAC 1551  
Qy 501 ArgGlyGlnThrSerProHisThrProAsnGluLysPheTyrGlyValThrValPheLys 520  
Db 1552 CGGGGTGAGACCTCCCCACATACCTCCCAATGAAAGTTCATGCTGCTCAAA 1611  
Qy 521 AlaLeuLysLeuGlyGlnGluGlyLysValProLeuGlnSerAlaHisLeuTyrTyrAsn 540  
Db 1612 GCCCTCAAGCTGGGGCAAGGCAAGTTCCTCTGCGAGAGTGCCACCTGTACTCAAC 1671  
Qy 541 ValThrGluLysValArgArgIleMetGluSerTyrPheArgLeuAspThrProLeuTyr 560  
Db 1672 GTGACGAGAGAAAGTGGCGGCATCATGAGTCTCTACTTCGCGCTGGATACGCCCTCTAC 1731  
Qy 561 PheSerTyrSerHisLeuValCysArgThrAlaIleGluValGlnAlaGluArgLys 580  
Db 1732 TTTTCTTACTCTCATCTGTTGTGCGCAGTCCCATCGAAGGTCCAGGAGAGGAG 1791  
Qy 581 AspAspSerHisProValHisValAspAsnCysIleLeuAsnAlaGluThrLeuValCys 600  
Db 1792 GATGATAGTATCCAGTCCACGTGACCACTGATGCTGTAATGCCGAGACCTCGTGT 1851  
Qy 601 ValLysGluProProAlaTyrThrPheArgAspTyrSerAlaIleLeuTyrLeuAsnGly 620  
Db 1852 GTCAAGAGCCCCAGGCTTACACCTTCGCGACTACAGCGCCATCTTTTACCTAAATGGG 1911  
Qy 621 AspPheAspGlyGlyAsnPheTyrPheThrGluLeuAspAlaLysThrValThrAlaGlu 640  
Db 1912 GACTTCGATGCGGAAACTTTTATTTCACTGAACCTGATGCCAAGACCGTGACGGCAGAG 1971  
Qy 641 ValGlnProGlnCysGlyArgAlaValGlyPheSerSerGlyThrGluAsnProHisGly 660  
Db 1972 GTGCAGGCTGTGAGAGAGCGGTGGGATTCCTTTCAAGCACTGAAACCCACATGGA 2031  
Qy 661 ValLysAlaValThrArgGlyGlnArgCysAlaIleAlaLeuTyrPheThrLeuAspPro 680  
Db 2032 GTGAGGCTGTACACGGGGGACGGGTGTGCCATCGCCCTGTGGTTCACTGGACCT 2091  
Qy 681 ArgHisSerGluArgAspArgValGlnAlaAspAspLeuValLysMetLeuPheSerPro 700  
Db 2092 CGACACAGCGAGCGGAGCAGGGTGCAGGAGATGACCTGGTGAAGATGCTCTTCAGCCCA 2151  
Qy 701 GluGluMetAspLeuSerGlnGluProLeuAspAlaGlnGlnGlyProProGluPro 720  
Db 2152 GAAGAGATGGACCTCTCCAGAGACGAGCCCTGATGCCAGAGCGGCCCCCGAACCT 2211  
Qy 721 AlaGlnGluSerLeuSerGlySerGluSerLysProLysAspGluLeu 736  
Db 2212 GCACAAGAGTCTCTCTCAGGAGTGAATCGAAGCCCAAGATGAGCTA 2259

RESULT 2  
AX136191  
LOCUS AX136191 2563 bp DNA linear PAT 30-MAY-2001  
DEFINITION Sequence 113 from Patent EP1067182.  
ACCESSION AX136191  
VERSION AX136191.1 GI:14272599  
KEYWORDS  
SOURCE Homo sapiens (human)  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE  
AUTHORS Ota,T., Isogai,T., Nishikawa,T., Kawai,Y., Sugiyama,T. and Hayashi,K.  
TITLE Secretory protein or membrane protein  
JOURNAL Patent: EP 1067182-A 113 10-JAN-2001;  
FEATURES Helix Research Institute (JP)  
source Location/Qualifiers  
1. .2563  
/organism="Homo sapiens"  
/mol\_type="unassigned DNA"  
/db\_xref="taxon:9606"  
31. .2241  
/note="unnamed protein product"  
/codon\_start=1  
/protein\_id="CAC39738.1"  
/db\_xref="GI:14272600"  
/db\_xref="REMBL:CAC39738"  
/translation="MAVRALKLLITLLAVAAASQAQAEVSEAGWGVTPDLLFAGBTA  
AYARDPQVVLSEMERALRRLRQTCADFPWELDPDWPSPQAQSGAA  
ALRDSFFGGLLRRAACLRCLPPAAHSLSEMELEFRKESPYNYLQVAFYKMKLE  
KAVAAAHFTFVGNPEHMEMQNLIDYOTMSGVKGADFKDLETQPMQFRGLGRLYSE  
NPOQAVPHEBAALQEQYFVAYEECRPSHYNVQFAYVNIQNTQAVCACTYLLPFPND  
NCKONCVTELAHSPREKPFEDFLPNSHYNVQFAYVNIQNTQAVCACTYLLPFPND  
EVMNONLAYAAMLGESETRSIGPRESAKEYRQSLLEKELLFFAYDVEGIPFVDPDS  
WTPBEVTPRIQERKOKSERATAVISQISGNLMEIEITLVEEKIKESLDVSRLLREGG  
PLLYEGISLTMNSKLLNGSRVMDGVISDHECOELQRLTVAATSGGIGYGTSPHT  
PNEKEYVTVPKALGQGVLPQSAHLYNTEKVRIMESFRLLTPILYFSYSHL  
VCRTEIEVQAEKDDSHPVNDNCILNAETLVCVKPPAYTFDYSLAILNGDFDG  
GNFYPTLDAKTVAEVOQCGRAVGFSSGTENPHGVKAVTRGQSCALALWFLDPRH  
SERDRVQADDLVKMLFSPENDLSQEQPLDAQQGPPEPAQESLSGSESKPDEL"

## ORIGIN

Alignment Scores:  
Pred. No.: 1.87e-293 Length: 2563  
Score: 3846.00 Matches: 733  
Percent Similarity: 99.59% Conservative: 0  
Best Local Similarity: 99.59% Mismatches: 3  
Query Match: 99.38% Indels: 0  
DB: 6 Gaps: 0

US-10-045-815-4 (1-736) x AX136191 (1-2563)

Qy 1 MetAlaValArgAlaLeuLysLeuLeuThrThrLeuLeuAlaValAlaAlaAlaSer 20  
Db 31 ATGGCGGTACGGCGGTTTGAAGTCTGTGACCACTGTGCTGGCTGTGCTGGCTGCCCTCC 90  
Qy 21 GlnAlaGluValGluSerGluAlaGlyTyrGlyMetValThrProAspLeuLeuPheAla 40  
Db 91 CAAGCCGAGTCCAGTCCGAGGAGGATGGGCGATGGTGAACCTGATCTGCTCTTCGCC 150  
Qy 41 GluGlyThrAlaAlaTyrAlaArgGlyAspTyrProGlyValValLeuSerMetGluArg 60  
Db 151 GAGGGACCGACGCTACCGCGCGGGGACTGGCCGGGGTGGTCTGAGCATGGAAACGG 210  
Qy 61 AlaLeuArgSerArgAlaAlaLeuArgAlaLeuArgLeuArgCysArgThrGlnCysAla 80  
Db 211 GCGCTGCGTCCCGGGCAGCCCTCCGCGCCCTTCGCTGCGCTGCCGACCCAGTGTGCC 270  
Qy 81 AlaAspPheProTyrProTyrProTyrProTyrProTyrProTyrProTyrProTyrProTyr 100  
Db 271 GCGGACTTCGCTGGGAGCTGACCCGCGCTGGTCCCCCAGCCCGCCCGCCCGCTCGGGC 330  
Qy 101 AlaGlyAlaLeuArgAspLeuSerPhePheGlyGlyLeuLeuArgArgAlaAlaCysLeu 120  
Db 331 GCGCGCGCTCGCGGACCTGAGCTTCTTCGGGGGCGCTTCTGCGTCCGCTGCTGCTGCTG 390  
Qy 121 ArgArgCysLeuGlyProProAlaAlaHisSerLeuSerGluGluMetGluLeuGluPhe 140  
Db 391 CGCGCTGCTCGGGCGCGCGCCCACTCGCTCAGCGAAGAGATGAGGTGAGGTTC 450  
Qy 141 ArgGlyArgSerProTyrAsnTyrLeuGlnValAlaTyrPheLysTyrLeuLysLeuGlu 160  
Db 451 CGCAAGCGAGCCCTTACAACTACTGCGAGGTGCGCTTCAAGATCAACAGTTGGAG 510

QY 161 LysAlaValAlaAlaAlaHisThrPhePheValGlyAsnProGluHisMetGluMetGln 180  
DB 511 AAAGTGTGTGTCGAGCAACACACCTTCTCGTGGGCAATCCTGAGCACATGAAATGCAG 570  
QY 181 GlnAsnLeuAspTyrTyrGlnThrMetSerGlyValLysGluAlaAspPheLysAspLeu 200  
DB 571 CAGAACCTAGACTATTACCAACCACTGCTGGAGTGAAGAGGCGGACTTCAAGGATCTT 630  
QY 201 GluThrGlnProHisMetGlnGluPheArgLeuGlyValArgLeuTyrSerGluGlnGln 220  
DB 631 GAGACTCAACCCCATATGCAAGAAATTCGACTGGGAGTGCAGACTCTACTCAGAGGAAACAG 690  
QY 221 ProGlnGluAlaValProHisLeuGluAlaAlaLeuGlnGluTyrPheValAlaTyrGlu 240  
DB 691 CCACAGGAGAGCTGTGCCCACTAGAGGCGGCTGCAAGAAATACTTTGTGGCTATGAG 750  
QY 241 GluCysArgAlaLeuGluGlyProTyrAspTyrAspGlyTyrAsnTyrLeuGluTyr 260  
DB 751 GAGTCCCGTGCCTCTGCGAAGGGCCCTATGACTACGATGGCTACAACACTACCTTGAATAC 810  
QY 261 AsnAlaAspLeuPheGlnAlaIleThrAspHisTyrIleGlnValLeuAsnCysLysGln 280  
DB 811 AACGCTGACCTCTCCAGGCCATCAGATCATATACATCCAGGTCCTCAACTGTGAAGCAG 870  
QY 281 AsnCysValThrGluLeuAlaSerHisProSerArgGluLysProPheGluAspPheLeu 300  
DB 871 AACTGTGTCCAGGAGCTTGTCCCAACCAAGTCGAGAGAAGCCCTTTGAAGACTTCCCTC 930  
QY 301 ProSerHisTyrAsnTyrLeuGlnPheAlaTyrTyrAsnIleGlyAsnTyrThrGlnAla 320  
DB 931 CCATCGCATTAATAATATCTGAGTTTGCCTACTATACATTGGGAATATACACAGGCT 990  
QY 321 GlyGluCysAlaLysThrTyrLeuLeuPhePheProAsnAspGluValMetAsnGlnAsn 340  
DB 991 GTTGAAATGTGCAAGACCTATCTTCTTCTCCCAATGACGAGGTGATGAACCAAAAT 1050  
QY 341 LeuAlaTyrTyrAlaAlaMetLeuGlyGluGluHisThrArgSerIleGlyProArgGlu 360  
DB 1051 TTGGCTATATTGACGATGATCTTGGAGAAGAACACACACAGATCCATCGGCCCGCTGAG 1110  
QY 361 SerAlaLysGluTyrArgGlnArgSerLeuLeuGluLysGluLeuPhePheAlaTyr 380  
DB 1111 AGTGCAAGGAGTACCGACGAGAGCTCTCTTCTCCCAATGACGAGGTGATGAACCAAAAT 1170  
QY 381 AspValPheGlyIleProPheValAspProAspSerTrpThrProGluGluValIlePro 400  
DB 1171 GATGTTTTTGGAAATTCCTTTTGTGATCCCGATTCATGGACTCCAGAGAAGTGTATCCC 1230  
QY 401 LysArgLeuGlnGluLysGlnLysSerGluArgGluThrAlaValArgIleSerGlnGlu 420  
DB 1231 AAGAGATTGCAAGAGAAACAGAAATGACAGACGGGAAACAGCCGTACGCATCTCCAGGAG 1290  
QY 421 IleGlyAsnLeuMetLysGluIleGluThrLeuValGluGluLysThrLysGluSerLeu 440  
DB 1291 ATTGGGAACCTTATGAAGAAATCGAGACCCCTTGTGGAAGAGAAGACCAAGGAGTCACTG 1350  
QY 441 AspValSerArgLeuThrArgGluGlyProLeuLeuTyrGluGlyIleSerLeuThr 460  
DB 1351 GATGTGACAGACTGACCGGGAAGGTGGCCCTCTGTATGAAGGCATCAGTCTCAAC 1410  
QY 461 MetAsnSerLysLeuLeuAsnGlyTyrGlnArgValValMetAspGlyValIleSerAsp 480  
DB 1411 ATGAATCCAACTCCTGAATGGTTCCAGCGGCTGTGATGAGCGGCTAATCTCTGAC 1470  
QY 481 HisGluCysGlnGluLeuGlnArgLeuThrAsnValAlaAlaThrSerGlyAspGlyTyr 500  
DB 1471 CACGAGTGTACGAGGCTGCAGAGACTGACCAATGTGGCAGCAACCTCAGGAGATGGCTAC 1530  
QY 501 ArgGlyGlnThrSerProHisThrProAsnGluLysPheTyrGlyValThrValPheLys 520  
DB 1531 CGGGGTGAGACTTCCCAACATCTCCCAATGAAGAAGTTCTATGTTGATCTGTTCAA 1590

QY 521 AlaLeuLysLeuGlyGlnGluGlyLysValProLeuGlnSerAlaHisLeuTyrTyrAsn 540  
DB 1591 GCCTCAAGCTGGGCAAGAGGCAAGTTCTCTGCGAGTGCACCTGTACTACTAAC 1650  
QY 541 ValThrGluLysValArgArgIleMetGluSerTyrPheArgLeuAspThrProLeuTyr 560  
DB 1651 GTGACGAGAGGTGGCGGCATCATGAGTCTCTACTTCCGCTCGATACGCCCTCTAC 1710  
QY 561 PheSerTyrSerHisLeuValCysArgThrAlaIleGluGluValGlnAlaGluArgLys 580  
DB 1711 TTTTCTACTCTCATCTGTGTGCGCACTGCGCATCGAAGAGTCCAGCGAGAGGAG 1770  
QY 581 AspAspSerHisProValHisValAspAsnCysIleLeuAsnAlaGluThrLeuValCys 600  
DB 1771 GATGATAGTATCCAGTCCACGTGGACAACCTGATCTCTGAATGCCGAGACCTCTGTGT 1830  
QY 601 ValLysGluProProAlaTyrThrPheArgAspTyrSerAlaIleLeuTyrLeuAsnGly 620  
DB 1831 GTCAAGAGAGCCCGACCTACACCTTCCGCACTACAGCGCATCTTACCTAAATGGG 1890  
QY 621 AspPheAspGlyGlyAsnPheTyrPheThrGluLeuAspAlaLysThrValThrAlaGlu 640  
DB 1891 GACTTCGATGGCGGAACCTTTTATTTCTGAACTGGATGCCAAGACCGTGACGCGCAG 1950  
QY 641 ValGlnProGlnCysGlyArgAlaValGlyPheSerSerGlyThrGluAsnProHisGly 660  
DB 1951 GTGACGCTCAGTGTGGAAGAGCCGTGGGATTTCTTTCAGGCACCTGAAACCCACATGA 2010  
QY 661 ValLysAlaValThrArgGlyGlnArgCysAlaIleAlaLeuTrpPheThrLeuAspPro 680  
DB 2011 GTGAAGGCTGTCCACAGGGGCGAGCGCTGTGCACTGCGCCCTGTGTTTCCCTCGACCCT 2070  
QY 681 ArgHisSerGluArgAspArgValGlnAlaAspAspLeuValLysMetLeuPheSerPro 700  
DB 2071 CGACACAGCGAGCGGACAGGCGTGCAGGACATGATGCTGGTGAAGATGCTCTTCAGCCCA 2130  
QY 701 GluGluMetAspLeuSerGlnGlnProLeuAspAlaGlnGlnGlyProProGluPro 720  
DB 2131 GAAGAGATGAGCTCTCCAGAGAGAGCCCTCGGATGCGCAGCGGCGCCCGGAGACCT 2190  
QY 721 AlaGlnGluSerLeuSerGlySerGluSerLysProLysAspGluLeu 736  
DB 2191 GCACAGAGTCTCTCTCAGGAGTGAATCGAAGCCCAAGATGAGCTA 2238  
RESULT 3  
BD123548 2563 bp DNA linear PAT 18-SEP-2002  
LOCUS  
DEFINITION Secretory protein or membrane protein.  
ACCESSION BD123548  
VERSION BD123548.1 GI:23218493  
KEYWORDS JP 2002017376-A/57.  
SOURCE Homo sapiens (human)  
ORGANISM Homo sapiens  
REFERENCE 1 (bases 1 to 2563)  
AUTHORS Ota,T., Isogai,T., Nishikawa,T., Kawai,Y., Sugiyama,T. and Hayashi,K.  
TITLES Secretory protein or membrane protein  
JOURNAL Patent: JP 2002017376-A 57 22-JAN-2002;  
HELIIX RESEARCH INSTITUTE  
COMMENT OS Homo sapiens (human)  
PN JP 2002017376-A/57  
PD 22-JAN-2002  
PF 07-JUL-2000 JP 2000253173  
PI TOSHIO OTA, TAKAO ISOGAI, TETSUO NISHIKAWA, YURI KAWAI, TOMOYASU  
PI SUGIYAMA,  
PI KOJI HAYASHI  
PC  
C12N15/09, C07K14/47, C07K16/18, C12N1/15, C12N1/19, C12N1/21, C12N5/ PC  
10, C12P21/02, C12Q1/68//C12P21/08, C12N15/00, C12N5/00 CC  
Secretory protein or membrane protein

FEATURES	source
FH Key	Location/Qualifiers
FT CDS	(31)..(2238).
	Location/Qualifiers
	1..2563
	/organism="Homo sapiens"
	/mol_type="genomic DNA"
	/db_xref="taxon:9606"
ORIGIN	
Alignment Scores:	
Pred. No.: 1-87e-293	Length: 2563
Score: 3846.00	Matches: 733
Percent Similarity: 99.59%	Conservative: 0
Best Local Similarity: 99.59%	Mismatches: 3
Query Match: 99.38%	Indels: 0
DB: 6	Gaps: 0
US-10-045-815-4 (1-736) x BD123548 (1-2563)	
QY 1	MetAlaValArgAlaLeuLysLeuThrThrLeuLeuAlaValValAlaAlaSer 20
Db	
31	ATGCGGGTACGGCGGTGAAGCTGTGCACCACACACTGCTGGCTGTCGTGGCGCTGCCCTCC 90
QY	21 GlnAlaGluValGluSerGluAlaGlyTrpGlyMetValThrProAspLeuPheAla 40
Db	
91	CAGCCGAGGTCCAGTCCGAGCAGGATGGGCGCATGGTGACCCGTGATCTGCTCTTCGCC 150
QY	41 GluGlyThrAlaAlaTyrralargGlyAspTrpProGlyValValLeuSerMetGluArg 60
Db	
151	GAGGGACCGCACCTACCGCGCGGGGACTCGCCCGGGGTGGTCTTAGCATGGAACGG 210
QY	61 AlaLeuArgSerArgAlaAlaLeuArgAlaLeuArgLeuArgCysArgThrGlnCysAla 80
Db	
211	CGCTCGCTCCCGGACACCTTCGGGCCCTTCGCTCGCTCGCGGACCCAGTGTGCC 270
QY	81 AlaAspPheProTrpGluLeuAspProAspTrpSerProAlaGlnAlaSerGly 100
Db	
271	GCGCACTTCGCTGGAGGTGACCCCGACTGGTCCCCAGCCCGCGCCGACGGCTCGGC 330
QY	101 AlaGlyAlaLeuArgAspLeuSerPhePheGlyGlyLeuLeuArgAtrcAlaAlaCysLeu 120
Db	
331	GCCGCGCCCTCGCGCACTGAGCTTTCTCGGGGGCCCTTCGCTCGCGCTGCTGCTG 390
QY	121 ArgArgCysLeuGlyProProAlaAlaHisSerLeuSerGluGluMetGluLeuGluPhe 140
Db	
391	CGCGCTGCCTCGGCGCGCGCGCCCACTCGCTCAGCGAAGAGATGGAGTTCGAGTTC 450
QY	141 ArgLysergSerProTyrrasnTyrrleuGlnValAlaTyrrPheLysIleAsnLysLeuGlu 160
Db	
451	CGAAAGCGGAGCCCCTACAACTACCTGCAAGTTCGCTTCTTCAAGATCAACAAGTGTGAG 510
QY	161 LysAlaValAlaAlaAlaHisThrPheValGlyAsnProGluHisMetGluMetGln 180
Db	
511	AARGCTGTCTCGACGACACACACTTCTTCGTGGGCAATCTCTGACCATGGAAATGCAG 570
QY	181 GlnAsnLeuAspTyrrTyrrGlnThrMetSerGlyValLysGluAlaAspPheLysAspLeu 200
Db	
571	CAGAACCTAGACTATTACCAAACCATGTCTGGAGTGAAGGCGGCACTTCAAGGATCTT 630
QY	201 GluThrGlnProHisMetGlnGluPheArgLeuGlyValArgLeuTyrrSerGluGluGln 220
Db	
631	GAGACTCAACCCCATATGAGAGATTTCGACTGGGAGTGCAGTCTACTCAGAGGAACAG 690
QY	221 ProGlnGluAlaValProHisLeuGluAlaAlaLeuGlnGluTyrrPheValAlaTyrrGlu 240
Db	
691	CCAAGAAAGTGTGCCCACTTAGAGCGGGGCTGCAAGAAATACTTTGTGGCTATGAG 750
QY	241 GluCysArgAlaLeuCysGluGlyProTyrrAspTyrrAspGlyTyrrAsnTyrrleuGluTyrr 260
Db	
751	GAGTGCCTGCTCGGAAGGGCCCTATGACTAGATGGCTACAACCTAGCTTGAAGTAC 810
QY	261 AsnAlaAspLeuPheGlnAlaAlaThrAspHisTyrrIleGlnValLeuAsnCysLysGln 280
Db	
811	AACGCTGACCTTCTCCAGGCCATCAGAGATCATATCATCCAGGTCTCTCACTGTAGCAG 870
QY	281 AsnCysValThrGluLeuAlaSerHisProSerArgGluLysProPheGluAspPheLeu 300
Db	
871	AACGTGTGCAGGAGCTTGTTCACCACCAAGTCGAGAGAAGCCCTTGAAGACTTCTCTC 930
QY	301 ProSerHisTyrrAsnTyrrleuGlnPheAlaTyrrAsnIleGlyAsnTyrrThrGlnAla 320
Db	
931	CCATCGCAATTAATATATCTGCAGTTTGCCTACTATAAACATTGGGAATTATATACAGGCT 990
QY	321 GlyGluCysAlaLysThrTyrrLeuLeuPhePheProAsnAspGluValMetAsnGlnAsn 340
Db	
991	GTTGAATGTGCAAGACCTATCTCTCTTCTCCCAATGACGAGGTGATGAACCAAAAT 1050
QY	341 LeuAlaTyrrTyrrAlaAlaMetLeuGlyGluGluHisThrArgSerIleGlyProArgGlu 360
Db	
1051	TTCGCTATTATGACGCTATGCTTTGGAGAAGAACACACCAAGATCCATCGCGCCCGCTGAG 1110
QY	361 SerAlaLysGluTyrrArgGlnArgSerLeuLeuGluLysGluLeuLeuPhePheAlaTyrr 380
Db	
1111	AGTGCAGGAGTACCGACGCGAAGCTACTTGGAAAAAGAACTGCTTTTTTCITCGCTTAT 1170
QY	381 AspValPheGlyIlePheProPheValAspProAspSerTrpThrProGluGluValIlePro 400
Db	
1171	GATGTTTTTGAATTTCCCTTTGTGATCCGATTCATGGATCTCAGAGAAGTGAITCCCT 1230
QY	401 LysArgLeuGlnGluLysGlnLysSerGluArgGluThrAlaValArgIleSerGlnGlu 420
Db	
1231	AAGAGATTGCAAGAGAACAAGAGTCAGAACGGGAACAGCCCTAGCATCTCCAGGAG 1290
QY	421 IleGlyAsnLeuMetLysGluIleGluThrLeuValGluGluLysThrLysGluSerLeu 440
Db	
1291	ATTGGGAACCTTATGAAGMAATCGAGACCTTGTGGAGAAGAACCAAGAGTCACTG 1350
QY	441 AspValSerArgLeuThrArgGluGlyGlyProLeuLeuTyrrGluGlyIleSerLeuThr 460
Db	
1351	GATGAGACAGACTGACCCGGAAGGTGGCCCCCTGCTGTATGAGGCAATCAGTCTCACC 1410
QY	461 MetAsnSerLysLeuLeuAsnGlyTyrrGlnArgValValMetAspGlyValIleSerAsp 480
Db	
1411	ATGAATCTCAAACTCCTGAATGTTCCAGCGGTGTGTATGGACGCGTAATCTCTGAC 1470
QY	481 HisGluCysGlnGluLeuGlnArgLeuThrAsnValAlaAlaThrSerGlyAspGlyTyrr 500
Db	
1471	CACGAGTGTGAGGAGCTGCAGAGACTGACCAATGTGGCAGCAACCTCAGAGATGGCTAC 1530
QY	501 ArgGlyGlnThrSerProHisThrProAsnGluLysPheTyrrGlyValThrValPheLys 520
Db	
1531	CGGGTCAAGACTCCCACTACATCTCCCAATGAAGAAGTCTATGGTGTCACTGTCTTCAA 1590
QY	521 AlaLeuLysLeuGlyGlnGluLysValProLeuGlnSerAlaHisLeuTyrrTyrrAsn 540
Db	
1591	GCCCTCAAGCTGGGCAAGAGCAAGTTCTCTGACAGAGTGCACCTGTACTACAAC 1650
QY	541 ValThrGluLysValArgAlaIleMetGluSerTyrrPheArgLeuAspThrProLeuTyrr 560
Db	
1651	GTGAGGAAAGAGGTGCGCGCATCATGGAGTCTTACTTCGCCCTGGATACGCCCTCTAC 1710
QY	561 PheSerTyrrSerHisLeuValCysArgThrAlaIleGluGluValGlnAlaGluArgLys 580
Db	
1711	TTTTCTACTCTCATCTGCTGTCGCGCACTGCCATCGAAGAGTTCACCTGTACTACAAC 1770
QY	581 AspAspSerHisProValHisValAspAsnCysIleLeuAsnAlaGluThrLeuValCys 600
Db	
1771	GATGATAGTCA





811 AACGTCACCTCTTCAGGCCATCAAGATCAATCATCAGGTCCTCAACTGTAAAGCAG 870  
Qy AsnCysValThrGluLeuAlaSerHisProSerArgGluLysProPheGluAspPheLeu 300  
Db AACTGTGTACAGGAGCTTGCTTCCACCAAGTCGAGAGAGCCCTTTGAAGACTTCCTC 930  
Qy ProSerHisTyrAsnTyrLeuGlnPheAlaTyrTyrAsnIleGlyAsnTyrThrGlnAla 320  
Db CCATCGCATTAATATATCTGCGAGTTTGCTTACTATAACATTGGGAATATATACAGGCT 990  
Qy 321 GlyGluCysAlaLysThrTyrLeuLeuPhePheProAsnAspGluValMetAsnGlnAsn 340  
Db 991 GTTGAATGTCGAAGCACTATCTTCTCTTCTTCCCAATGACAGGTGATGAACCAAT 1050  
Qy 341 LeuAlaTyrTyrAlaAlaMetLeuGlyGluGluHisThrArgSerIleGlyProArgGlu 360  
Db 1051 TTGGCTATTATGAGCTATGCTTGGAGAAGAACACACAGATCCATCGCCGCCCTGAG 1110  
Qy 361 SerAlaLysGluTyrArgGlnArgSerLeuLeuGluLysGluLeuPhePheAlaTyr 380  
Db 1111 AGTGCCAAGGAGTACCGACAGCAAGCTTACTGGAAGAAAGAACTGCTTCTTTCGCTTAT 1170  
Qy 381 AspValPheGlyIleProPheValAspProAspSerTrpThrProGluGluValIlePro 400  
Db 1171 GATGTTTTTGAATTCCTTTGTGGATCCGATTTCGACTCCAGAGAGATGATTCCT 1230  
Qy 401 LysArgLeuGlnGluLysGlnLysSerGluArgGluThrAlaValArgIleSerGlnGlu 420  
Db 1231 AAGAGATTTGCAAGAGAACAAGAGTCAGAACCGGAACACCGCTACGCTATCTCCAGGAG 1290  
Qy 421 IleGlyAsnLeuMetLysGluIleGluThrLeuValGluGluLysThrLysGluSerLeu 440  
Db 1291 ATTGGGACCTTATGAAGGAATTCGAGACCTTGTGGAGAGAACCAAGAGTCACTG 1350  
Qy 441 AspValSerArgLeuThrArgGluGlyGlyProLeuLeuTyrGluGlyIleSerLeuThr 460  
Db 1351 GATGTGAGCAGACTGACCCGGGAAGGTGGCCCTGCTGTATGAAGGCATCACTTCACC 1410  
Qy 461 MetAsnSerLysLeuLeuAsnGlyTyrGlnArgValValMetAspGlyValIleSerAsp 480  
Db 1411 ATGAACCTCAAACTCTCGAATGTTCCAGCGGTGTGTGAGACGGCGTATCTCTGAC 1470  
Qy 481 HisGluCysGlnGluLeuGlnArgLeuThrAsnValAlaAlaThrSerGlyAspGlyTyr 500  
Db 1471 CACGAGTGTGAGGAGCTGCAGAGACTGACCAATGTGGCAACCTCAGGAGATGCTAC 1530  
Qy 501 ArgGlyGlnThrSerProHisThrProAsnGluLysPheTyrGlyValThrValPheLys 520  
Db 1531 CGGGGTGAGACCTCCACACATCTCCCAATGAAAGTTCTATGGTGTCACTGTCTTCAA 1590  
Qy 521 AlaLeuLysLeuGlyGlnGluGlyLysValProLeuGlnSerAlaHisLeuTyrTyrAsn 540  
Db 1591 GCCCTCAAGCTGGGGCAAGAGCAAGATTCTCTGAGAGTGCACACCTGTACTACAC 1650  
Qy 541 ValThrGluLysValArgGlyMetGluSerTyrPheArgLeuAspThrProLeuTyr 560  
Db 1651 GTGACGGAGAAGGTGCGCGCATCATGGAGTCTCTACTTCGCGCTGGATACGCCCTCTAC 1710  
Qy 561 PheSerTyrSerHisLeuValCysArgThrAlaIleGluGluValGlnAlaGluArgLys 580  
Db 1711 TTTTCTCTACTCTCATCTGTTGGTGGCGACCTGCCATCGAAGGTCCAGCGAGAGGAG 1770  
Qy 581 AspAspSerHisProValHisValAspAsnCysIleLeuAsnAlaGluThrLeuValCys 600  
Db 1771 GATGATAGTCATCCAGTCCACGTCGACATCTGATGCACTCTGATGCCAGACCCCTCGTGT 1830  
Qy 601 ValLysGluProProAlaTyrThrPheArgAspTyrSerAlaIleLeuTyrLeuAsnGly 620  
Db 1831 GTCAAGAGAGCCCCAGGCTTACCTTCGCGACTACAGCGCATCTTACCTTAATATGGG 1890  
Qy 621 AspPheAspGlyGlyAsnPheTyrPheThrGluLeuAspAlaLysThrValThrAlaGlu 640  
Db 1891 GACTTCGATGCGGGAACCTTTATTTTCACTGAAGTGGATGCCAAGACCGTGACGGCAGAG 1950

Qy 641 ValGlnProGlnCysGlyArgAlaValGlyPheSerSerGlyThrGluAsnProHisGly 660  
Db 1951 GTGCAGCCTCAGTGTGGAGAGCCGTGGGATTTCTTTTTCAGCACTGAAACCCCATGGA 2010  
Qy 661 ValLysAlaValThrArgGlyGlnArgCysAlaIleAlaLeuTyrPheThrLeuAspPro 680  
Db 2011 GTGAAGGCTGTACACAGGGGGCAGGGCTGTGCCATCGCCCTGTGGTTCACCTTGACCT 2070  
Qy 681 ArgHisSerGluArgAspArgValGlnAlaAspAspLeuValLysMetLeuPheSerPro 700  
Db 2071 CGACACAGCGAGCGGACAGGGTGCAGGAGATGACCTGGTGAAGATGCTCTTCAGCCCA 2130  
Qy 701 GluGluMetAspLeuSerGlnGluProLeuAspAlaGlnGlnGlyProProGluPro 720  
Db 2131 GAAGAGATGGACCTCTCCAGAGAGCCCTTGGATGCCAGCAGGGCCCCCCCCAACCT 2190  
Qy 721 AlaGlnGluSerLeuSerGlySerGluSerLysProLysAspGluLeu 736  
Db 2191 GCACAGAGATCTCTCTCAGGAGTGAATCGAAGCCCAAGGATGAGCTA 2238

## RESULT 5

AX704765 2593 bp DNA linear PAT 04-APR-2003  
LOCUS Sequence 69 from Patent WO0202634.  
DEFINITION AX704765

ACCESSION AX704765.1 GI:29561431

VERSION Homo sapiens (human)

KEYWORDS Homo sapiens

SOURCE Homo sapiens

ORGANISM Homo sapiens

REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

AUTHORS Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

1 Gururajan, R., Hafalia, A.J., Kallick, D.A., Patterson, C., Azimzai, Y.,

Khan, F.A., Xu, Y., Yao, M.G., Yue, H., Au-Young, J., Batra, S., Y.T.,

Baughn, M.R., Borowsky, E.A., Lo, I.F., Lu, D.A., Lu, X., Tang, Y.T.,

Yang, J., Zingler, X.A., Deleane, A.M., Gietzen, K., Marcus, G.A.,

Nguyen, D.B., Policky, J.L., Ramkumar, J., Thangavelu, K., Wallia, N.K.

and Warren, B.A.

Human extracellular matrix and cell adhesion polypeptides

PATENT: WO 0202634-A 69 10-JAN-2002;

INCITE GENOMICS INC. (US)

FEATURES Location/Qualifiers

1..2593

/organism="Homo sapiens"

/mol\_type="unassigned DNA"

/db\_xref="taxon:9606"

/note="incyte ID No: 6780147CB1"

ORIGIN

Alignment Scores:

Pred. No.: 1.89e-293 Length: 2583

Score: 3846.00 Matches: 733

Percent Similarity: 99.59% Conservative: 0

Best Local Similarity: 99.58% Mismatches: 3

Query Match: 99.38% Indels: 0

DB: Gaps: 6

US-10-045-815-4 (1-736) x AX704765 (1-2583)

Qy 1 MetAlaValArgAlaLeuLysLeuThrThrLeuAlaValAlaAlaSer 20

Db 42 ATGGCGGTACGCGGTGTGAAGCTGTGACCACTGTGGCTGTGTGGCCCTGCTCC 101

Qy 21 GlnAlaGluValGluSerGluAlaGlyTrpGlyMetValThrProAspLeuPheAla 40

Db 102 CAAGCCGAGTGTGAGTCCGAGGAGGATGGGATGTGTGACGCTGATCTCTTCGCC 161

Qy 41 GluGlyThrAlaAlaTyrAlaArgGlyAspTrpProGlyValValLeuSerMetGluArg 60

Db 162 GAGGGGACCGCAGCTTACCGCGCGGGGACTGGCCGGGGTGGTCTGAGCATGGAACGG 221

Qy 61 AlaLeuArgSerArgAlaAlaLeuArgAlaLeuArgLeuArgCysArgThrGlnCysAla 80

Db 222 GGGCTGCGCTCCGGGCGAGCCCTCGCGCCCTTCGCTCGCGTCCGACCCAGTGTGCC 281  
Qy 81 AlaAspPheProTribLeuAspProAspTrpSerProSerProAlaGlnAlaSerGly 100  
Db 282 GCCGACTTCCCGTGGAGCTGAGCCCGGAGTGGTCCCGGAGGCGCCGAGGCTCGGGC 341  
Qy 101 AlaGlyAlaLeuArgAspLeuSerPhePheGlyGlyLeuLeuArgAtgAlaAlaCysLeu 120  
Db 342 GCCGCGCCCTCGCGAGCTAGACTTCTTCGCGGCGCTTCGCGTGGCTGCGCTG 401  
Qy 121 ArgArgCysLeuGlyProProAlaAlaHisSerLeuSerGluGluMetGluLeuGluPhe 140  
Db 402 CGCCGCTGCTCGGCGCCCGCGCCCGCCACTCGCTCAGCAGAGATGGAGCTGGAGTTC 461  
Qy 141 ArgLysArgSerProGlyAsnTyrLeuGlnValAlaTyrPheLysIleAsnLysLeuGlu 160  
Db 462 CGCAAGCGGAGCCCTACACTACCTGCGAGTTCGCTACTTCAAGATCAACAGTTGGAG 521  
Qy 161 LysAlaValAlaAlaAlaHisThrPhePheValGlyAsnProGluHisMetGluMetGln 180  
Db 522 AAAGCTGTGTGTCAGCACACACCTTCTTCGTGGGCAATCTCGAGCACATGGAATGCAG 581  
Qy 181 GlnAsnLeuAspTyrTyrGlnThrMetSerGlyValLysGluAlaAspPheLysAspLeu 200  
Db 582 CAGAACCTAGACTATTACCAACCATGTCGTGAGTGAAGAGGCGGACTTCAAGGATCTT 641  
Qy 201 GluThrGlnProHisMetGlnGluPheArgLeuGlyValArgLeuTyrSerGluGluGln 220  
Db 642 GAGACTCAACCCCATATGCAAGAAATTCGACTGGAGGTGCGACTCTACTCAGAGCAACAG 701  
Qy 221 ProGlnGluAlaValProHisGluAlaAlaLeuGlnGluTyrPheValAlaTyrGlu 240  
Db 702 CCACAGGAAGCTGTGCCCCACTAGAGCGCGCTGCAAGAAATACTTTGTGGCCATGAG 761  
Qy 241 GluCysArgAlaLeuCysGluGlyProTyrAspTyrAspGlyTyrAsnTyrLeuGluTyr 260  
Db 762 GAGTCCGCTGCGCTCTCGAAGGCGCCCTATGACTACGATGCGTACAACTACCTTGAGTAC 821  
Qy 261 AsnAlaAspLeuPheGlnAlaIleThrAspHisTyrIleGlnValLeuAsnCysLysGln 280  
Db 822 AACGCTGACTCTTCCAGGCGCATCACAGATCATTACATCCAGGTGCTTCAACTGTAAAGCAG 881  
Qy 281 AsnCysValThrGluLeuAlaSerHisProSerArgGluLysProPheGluAspPheLeu 300  
Db 882 AACTGTGTACGGAGCTGTCTCCCAAGTCCAGAGTCCAGAGCCCTTGAAGACTTCTCTC 941  
Qy 301 ProSerHisTyrAsnTyrLeuGlnPheAlaTyrTyrAsnIleGlyAsnTyrThrGlnAla 320  
Db 942 CCATCGCATTAATAATTATCTGAGTTTGCCTTACTATAACATTGGGAATTTATACACAGGCT 1001  
Qy 321 GlyGluCysAlaLysThrTyrLeuLeuPhePheProAsnAspGluValMetAsnGlnAsn 340  
Db 1002 GTTGAAATGTGCAAGACTATCTTCTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCT 1061  
Qy 341 LeuAlaTyrTyrAlaAlaMetLeuGlyGluGluHisThrArgSerIleGlyProArgGlu 360  
Db 1062 TTGGCTTATTATGAGCTATGCTTGGAGAGAAACACACACAGATCCATCGGCCCGCGTGAG 1121  
Qy 361 SerAlaLysGluTyrArgGlnArgSerLeuLeuGluLysGluLeuLeuPhePheAlaTyr 380  
Db 1122 AGTCCCAAGGAGTACCCAGCAGCAGGAGCTTACTGGAAGAAAGAACTGCTTTCTTCGCTTAT 1181  
Qy 381 AspValPheGlyIleProPheValAspProAspSerTrpThrProGluGluValIlePro 400  
Db 1182 GATGTTTTTGGAAATTCCTTTGTGATCCGGATTCATGGACTCCAGAGAAGTGAATCC 1241  
Qy 401 LysArgLeuGlnGluLysGlnLysSerGluArgGluThrAlaValArgIleSerGlnGlu 420  
Db 1242 AAGAGATTGCAAGAGAAACAGAGTCAAGCGGAAACAGCCGTACGCATCTCCCGAGGAG 1301  
Qy 421 IleGlyAsnLeuMetLysGluIleGluThrLeuValGluGluLysThrLysGluSerLeu 440

Db 1302 ATTGGGAACCTTATGAAGGAAATCGAGACCCCTTGTGGAGAGAGAACCAAGAGTCACTG 1361  
Qy 441 AspValSerArgLeuThrArgGluGlyGlyProLeuLeuTyrGluGlyLysLeuThr 460  
Db 1362 GATGTAGCAGACTGACCCGGAAGTGGCCCTCTGCTGTATGAAGGATCATGTTCTCAC 1421  
Qy 461 MetAsnSerLysLeuLeuAsnGlyTyrGlnArgValMetAspGlyValIleSerAsp 480  
Db 1422 ATGAACCTCCAAATCTCTGAATGGTCCCGAGCGGGTGGTATGAGCGGTATCTCTGAC 1481  
Qy 481 HisGluCysGlnGluLeuGlnArgLeuThrAsnValAlaAlaThrSerGlyAspGlyTyr 500  
Db 1482 CAGAGTGTGAGAGCTGAGAGACTCCCAATGTGGAGCAACCTCAGGAGATGGCTAC 1541  
Qy 501 ArgGlyGlnThrSerProHisThrProAsnGluLysPheTyrGlyValThrValPheLys 520  
Db 1542 CGGGGTGAGACCTCCCACTACTCCCAATGAAAGTTCATGGTGTCTGCTCTCAA 1601  
Qy 521 AlaLeuLysLeuGlyGlnGluLysValProLeuGlnSerAlaHisLeuTyrTyrAsn 540  
Db 1602 GCCCTCAAGCTGGGGCAGAGAGGCAAGTTCCTCTGAGAGTGCACCTGTACTACAAC 1661  
Qy 541 ValThrGluLysValArgArgIleMetGluSerTyrPheArgLeuAspThrProLeuTyr 560  
Db 1662 GTGACGAGAGAGTGGGGCGCATCATGAGTCTTACTTCCGCTCGATACGCCCTCTTAC 1721  
Qy 561 PheSerTyrSerHisLeuValCysArgThrAlaIleGluGluValGlnAlaGluArgLys 580  
Db 1722 TTTTCTACTCTCATCTGGTGTGCGCACTGCAATCGAAGAGTCCAGGAGAGAGGAG 1781  
Qy 581 AspAspSerHisProValHisValAspAsnCysIleLeuAsnAlaGluThrLeuValCys 600  
Db 1782 GATGATAGTATCCAGTCCAGTGCAGCACTGATCTGATGCGAGACCTCTGCTGTGT 1841  
Qy 601 ValLysGluProProAlaTyrThrPheArgAspTyrSerAlaIleLeuTyrLeuAsnGly 620  
Db 1842 GTCAAGAGAGCCCGGCTTACACTTCCGCGACTACAGCGCATCTTTTACCTAAATGGG 1901  
Qy 621 AspPheAspGlyGlyAsnPheTyrPheThrGluLeuAspAlaLysThrValThrAlaGlu 640  
Db 1902 GACTTCGATGGCGAAACTTTTATTCTACTGAATCGGATGCCAAGACCGTGACGCGAG 1961  
Qy 641 ValGlnProGlnCysGlyArgAlaValGlyPheSerSerGlyThrGluAsnProHisGly 660  
Db 1962 GTGAGCCTCAGTGTGGAAGAGCGCTGGGATTCCTTCAGGCACTGAAACCCACATGA 2021  
Qy 661 ValLysAlaValThrArgGlyGlnArgCysAlaIleAlaLeuTrpPheThrLeuAspPro 680  
Db 2022 GTGAAGGCTGTCAACGAGGCGAGCGCTGTGCCATCGCCCTGTGGTTTCACTGACCCCT 2081  
Qy 681 ArgHisSerGluArgAspArgValGlnAlaAspAspLeuValLysMetLeuPheSerPro 700  
Db 2082 CGACACGAGGAGCGGAGCAGGCTGAGCAGCATGACCTGTGTGAAGATGCTCTTCAGGCCA 2141  
Qy 701 GluGluMetAspLeuSerGlnGluGlnProLeuAspAlaGlnGlnGlyProProGluPro 720  
Db 2142 GAAGAGATGGACCTCTCCAGAGCAGCCCTCGATGTCAGCAGGCGGCCCCCGGAACT 2201  
Qy 721 AlaGlnGluSerLeuSerGlySerCysLysProLysAspGluLeu 736  
Db 2202 GCACAAGAGTCTCTCTCAGGAGTGAATCGAAGCCCAAGGATGAGCTA 2249

RESULT 6  
AX877288  
LOCUS  
DEFINITION  
Sequence 12193 from Patent EPI074617.  
AX877288  
ACCESSION  
AX877288.1 GI:40032024  
VERSION  
KEYWORDS  
SOURCE  
ORGANISM  
Homo sapiens (human)  
Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

CDS 53.2263

```

Db      1613 GCCCTCAAGCTGGGCAAGGCAAGAAAGTTCTCTGCAGAGTGCCACCTGACTACAA 1672
Qy      541 ValThrGluValArgArgIleVerGluSerTyrPheArgLeuAspThrProLeuTyr 560
Db      1673 GTGACGGAGAGTGGGGCATGATGAGTCTTACTTCCGCTGGATACGCCCTCTTAC 1732
Qy      561 PheSerTyrSerHisLeuValCysArgThrAlaIleGluValGlnAlaGluArgLys 580
Db      1733 TTTTCTACTCTCATCTGGTGGCGCATGCGCATCGAAGAGGTCCAGGCGAGAGGAAG 1792
Qy      581 AspAspSerHisProValHisValAspAsnCysIleLeuAsnAlaGluThrLeuValCys 600
Db      1793 GATGATAGTCATCCAGTCCAGTGGACAACTGCATCTCTGAATGGGAGACCTCGTGTGT 1852
Qy      601 ValLysGluProProAlaTyrThrPheArgAspTyrSerAlaIleLeuTyrLeuAsnGly 620
Db      1853 GTCAAGAGGCCCCAGCCTACACCTTCGCGACTACAGCGCATCTTTACTTAATGGG 1912
Qy      621 AspPheAspGlyGlyAsnPheTyrPheThrGluLeuAspAlaLysThrValThrAlaGlu 640
Db      1913 GACTTCGATGGCGGAACCTTTTATTTCACTGAATGGATGCCAAGACCGTGCAGCGCAGAG 1972
Qy      641 ValGlnProGlnCysGlyArgAlaValGlyPheSerSerGlyThrGluAsnProHisGly 660
Db      1973 GTGCAAGCTCAGTGTGGAGAGCCGTGGATCTCTTCAGGCATCTGAACCCACATGGA 2032
Qy      661 ValLysAlaValThrArgGlyGlnArgCysAlaIleAlaLeuTyrPheThrLeuAspPro 680
Db      2033 GTGAAGGCTGTCAACAGGGGCGAGCGCTGTGCCATGCGCCCTGTGGTTCACCCCTGAC 2092
Qy      681 ArgHisSerGluArgAspArgValGlnAlaAspAspLeuValLysMetLeuPheSerPro 700
Db      2093 CGACACAGCGAGCGGAGCAGGGTGCAGGCAGATGACCTGTGTGAAGATGCTCTTCAGGCC 2152
Qy      701 GluGluMetAspLeuSerGlnGluGlnProLeuAspAlaGlnGlnGlyProProGluPro 720
Db      2153 GAAGAGATGGACCTCTCCAGAGAGCAGCCCTGGATGCCAGCAGGCGCCCGCCGACCT 2212
Qy      721 AlaGlnGluSerLeuSerGlySerGluSerLysProLysAspGluLeu 736
Db      2213 GCACAAGAGTCTCTCTCAGGAGTGAATCGAAGCCCAAGGATGAGCTA 2260

RESULT 7
LOCUS   BD156581 2585 bp DNA linear PAT 17-JAN-2003
DEFINITION  Primer for synthesizing full-length cDNA and use thereof.
ACCESSION  BD156581
VERSION    BD156581.1 GI:27862339
KEYWORDS   JP 2002191363-A/11424.
SOURCE     Homo sapiens (human)
ORGANISM   Homo sapiens
           Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
           Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE  1 (bases 1 to 2585)
AUTHORS   Ota,T., Isogai,T., Nishikawa,T., Hayashi,K., Saito,K., Yamamoto,J.,
           Ishii,S., Sugiyama,T., Wakamatsu,A., Nagai,K. and Otsuki,T.
TITLE     Primer for synthesizing full-length cDNA and use thereof
JOURNAL   Patent: JP 2002191363-A 11424 09-JUL-2002;
           HELIX RESEARCH INSTITUTE
COMMENT    OS Homo sapiens (human)
           PN JP 2002191363-A/11424
           PD 09-JUL-2002
           PF 28-JUL-2000 JP 2000280990
           PI TOSHIO OTA,TAKAO ISOGAI,TEISUO NISHIKAWA,KOJI HAYASHI,KAORU
           PI SAITO,
           PI JUNICHI YAMAMOTO,SHIZUKO ISHII,TOMOYASU SUGIYAMA,AI WAKAMATSU,
           PI KEIICHI NAGAI,TEISUJI OTSUKI
           PC C12N15/09,C07K14/47,C07K16/18,C12N1/15,C12N1/19,C12N1/21,C12N5/ PC
           10,
           PC C12P21/02,C12Q1/68//C12P21/08,G06F17/30,C12N15/00,C12N5/00 CC
           Primer for synthesizing full-length cDNA and use thereof FH key

```

```

FEATURES             FT  CDS           Location/Qualifiers
     source            1..2585          Location/Qualifiers
                        /organism="Homo sapiens"
                        /mol_type="genomic DNA"
                        /db_xref="taxon:9606"

ORIGIN
Alignment Scores:      1.89e-293      Length:      2585
Pred. No.:            3846.00         Matches:    733
Score:                99.59%          Conservative: 0
Percent Similarity:   99.59%          Mismatches:  3
Best Local Similarity: 99.59%          Indels:      0
Query Match:          99.38%          Gaps:        6
DB:
US-10-045-815-4 (1-736) x BD156581 (1-2585)

Qy      1 MetAlaValArgAlaLeuLysLeuThrThrLeuLeuAlaValAlaAlaLaSer 20
Db      53 ATGCGCGTACGGCGGTGAAGCTGCTGACCACTGCTGGCTGTGCTGCCCGCTGCTCC 112
Qy      21 GlnAlaGluValGluSerGluAlaGlyTyrGlyMetValThrProAspLeuLeuPheAla 40
Db      113 CAAAGCGAGGTGAGTCCGAGGAGGATGGGCGATGGTGACGCCGTGATCTGCTCTTCGCC 172
Qy      41 GluGlyThrAlaAlaTyrAlaArgGlyAspTyrProGlyValValLeuSerMetGluArg 60
Db      173 GAGGGACCGACGACCTACGCGCGCGGGAGCTGCGCCGGGGTGGTCTCTGAGCATGGAACGG 232
Qy      61 AlaLeuArgSerArgAlaAlaLeuArgAlaLeuArgLeuArgCysArgThrGlnCysAla 80
Db      233 GCGCTGCGCTCCCGGCGACCTCCGCGCCCTTCGCTCGCGTGGCGCCAGCCAGGTGCC 292
Qy      81 AlaAspPheProTyrGluLeuAspProAspTyrProSerProAlaGlnAlaSerGly 100
Db      293 GCGGACTTCCTCGTGGAGCTGGACCCCGACTGCTCCCGACCGCCGCGCCGCGCTCGGCG 352
Qy      101 AlaGlyAlaLeuArgAspLeuSerPhePheGlyGlyLeuLeuArgArgAlaAlaCysLeu 120
Db      353 GCGCGCGCCCTCGCGACCTGAGCTTCTTCGGGGGCTTCTGCGTGGCGCTGCGCTGCGCTG 412
Qy      121 ArgArgCysLeuGlyProProAlaAlaHisSerLeuSerGluLeuMetGluLeuGluPhe 140
Db      413 CGCCCGCTGCTCGGGCGCGCGCCCGCCCGCCACTCGCTCAGCGAAGAGATGGAGCTGGAGTTC 472
Qy      141 ArgLysArgSerProTyrAsnTyrLeuGlnValAlaTyrPheLysIleAsnLysLeuGlu 160
Db      473 CGCAAGCGGAGCGCCCTACAACTACCTGCGAGTGGCTACTTCAAGATCAACAAGTTGGAG 532
Qy      161 LysAlaValAlaAlaAlaHisThrPheValGlyAsnProGluHisMetGluMetGln 180
Db      533 AAAGCTGTGCTGCGAGCACACACCTTCTTCGGGGCAATCTCTGAGCATGGAATGGAG 592
Qy      181 GlnAsnLeuAspTyrTyrGlnThrMetSerGlyValLysGluAlaAspPheLysAspLeu 200
Db      593 CAGAACCTAGACTATTACCAACCATGCTGAGAGTGAAGGAGGCGGACTTCAAGGATCTT 652
Qy      201 GluThrGlnProHisMetGlnGluPheArgLeuGlyValArgLeuTyrSerGluGluGln 220
Db      653 GAGACTCAACCCCATATCAAGAAATTCGATGGGAGTGGGACTCTACTCAGAGGAACAG 712
Qy      221 ProGlnGluAlaValProHisLeuGluAlaAlaLeuGlnGluTyrPheValAlaTyrGlu 240
Db      713 CCACAGGAAGCTGTGCCCACTAGAGCGCGCTGCAAGAATACTTTGTGGCCCTATGAG 772
Qy      241 GluCysArgAlaLeuCysGluGlyProTyrAspTyrAspGlyTyrAsnTyrLeuGluTyr 260
Db      773 GAGTCCCGTGGCTCTCGGAAGGGGCCCTATGACTACGATGGCTCAACTACTCTTGAAGTAC 832
Qy      261 AsnAlaAspLeuPheGlnAlaIleThrAspHisTyrIleGlnValLeuAsnCysLysGln 280

```

833 AACGCTGACCTCTTCAGGCCATCAGCATCATTTACATCCAGGTCTCAACTGTAAACAG 892  
281 AsnCysValThrGluLeuAlaSerHisProSerArgGluLysProPheGluAspPheLeu 300  
893 AACTGTGTCAGGAGCTGCTTCCCAACCAAGTCGAGAGAGCCCTTTGAAGACTTCCTC 952  
301 ProSerHisTyrAsnTyrLeuGlnPheAlaTyrTyrAsnIleGlyAsnTyrThrGlnAla 320  
953 CCATCGCATTAATATATCTGCGAGTTGCTCTACTATAACATTGGGAATTATACACAGGCT 1012  
321 GlyGluCysAlaLysThrTyrLeuLeuPhePheProAsnAspGluValMetAsnGlnAsn 340  
1013 GTTGAATGTGCCAAGACTATCTTCTTCTCCCAATGACGAGGTGATGAACCAAAAT 1072  
341 LeuAlaTyrTyrAlaAlaMetLeuGlyGluGluHisThrArgSerIleGlyProArgGlu 360  
1073 TTGGCCATTAATGTCAGCTATGCTTGGAGAGACACACACAGATCCATCGCCCCCTGAG 1132  
361 SerAlaLysGluTyrArgGlnArgSerLeuLeuGluLysGluLeuLeuPheAlaTyr 380  
1133 AGTCCCAAGAGTACCCGACAGCAAGCTACTGCGAAAAGAACTGCTTTCTTCGCTTAT 1192  
381 AspValPheGlyIleProPheValAspProAspSerTyrThrProGluGluValIlePro 400  
1193 GATGTTTTTGGAAATTCCTTTTGGATCCGGATTCATGGACTCCAGAAGAGTATTCCTC 1252  
401 LysArgLeuGlnGluLysGlnLysSerGluArgGluThrAlaValArgIleSerGlnGlu 420  
1253 AAGAGATTGACAGAGAACAGAGTCAAGACGGGAAACACCGCTAGCATCTCCAGGAG 1312  
421 IleGlyAsnLeuMetLysGluIleGluThrLeuValGluLysThrLysGluSerLeu 440  
1313 ATTGGGAACCTTATGAAGAGAAATCGAGACCTTTGTGGAGAGAGACCAAGGAGTCACTG 1372  
441 AspValSerArgLeuThrArgGluGlyGlyProLeuLeuTyrGluGlyIleSerLeuThr 460  
1373 GATGTGACGACATGACCCGGGAGGTGGCCCCCTGCTGTATGAGGACATCATCTCAC 1432  
461 MetAsnSerLysLeuLeuAsnGlyTyrGlnArgValValMetAspGlyValIleSerAsp 480  
1433 ATGAACCTCAAACTCTCTGAATGTTTCCAGCGGTGTGTATGACGGCGTAATCTCTGAC 1492  
481 HisGluCysGlnGluLeuGlnArgLeuThrAsnValAlaAlaThrSerGlyAspGlyTyr 500  
1493 CACAGGTGTGAGGAGCTGACAGAGACTACCAATGTGCAGCAACCTCAGGAGATGGCTAC 1552  
501 ArgGlyGlnThrSerProHisThrProAsnGluLysPheTyrGlyValThrValPheLys 520  
1553 CGGGGTGAGACCTCCCCACATACCTCCCAATGAAAGTTCTATGTTGTCACGTCTTCAA 1612  
521 AlaLeuLysLeuGlyGlnGluGlyLysValProLeuGlnSerAlaHisLeuTyrTyrAsn 540  
1613 GCCCTCAAGCTGGGGCAAGAGGCAAAAGTTCTCTGCAGAGTGCCTGCTACTCAAC 1672  
541 ValThrGluLysValArgArgIleMetGluSerTyrPheArgLeuAspThrProLeuTyr 560  
1673 GTGACGGAGAGGTGGCGGCGATCATGGAGTCTCTCTCGGCTGTGATCGCCCTCTAC 1732  
561 PheSerTyrSerHisLeuValCysArgThrAlaIleGluGluValGlnAlaGluArgLys 580  
1733 TTTTCTCTACTCTCATCTGTTGTCGCACTGCCATCGAAGAGTCCAGGCGAGAGGAG 1792  
581 AspAspSerHisProValHisValAspAsnCysIleLeuAsnAlaGluThrLeuValCys 600  
1793 GATGATAGTCATCCAGTCCAGTGCACCTGCACTTCCTGATGTCGAGACCTCTGTTGT 1852  
601 ValLysGluProProAlaTyrThrPheArgAspTyrSerAlaIleLeuTyrLeuAsnGly 620  
1853 GTCAAGAGGCCCCAGCGCTACCTTCGCGACTACAGCGCATCTCTTACTTAATGGG 1912  
621 AspPheAspGlyGlyAsnPheTyrPheThrGluLeuAspAlaLysThrValThrAlaGlu 640  
1913 GACTTCGATGCGGAAACTTTTATTTCACTGAATGCTGCAAGCCGTCGAGCGGAGAG 1972

641 ValGlnProGlnCysGlyValArgAlaValGlyPheSerSerGlyThrGluAsnProHisGly 660  
1973 GTGACGCTCTAGTGTGGAGAGCGCTGGGATTCCTTCAGGACACTGAAACCCCATGGA 2032  
661 ValLysAlaValThrArgGlyGlnArgCysAlaIleAlaLeuTrpPheThrLeuAspPro 680  
2033 GTGAAGGCTGTCCACAGGGGCGAGCGCTGTGCCATCGCCCTGTGTTTCACTCGACCT 2092  
681 ArgHisSerGluArgAspValGlnAlaAspAspLeuValLysMetLeuPheSerPro 700  
2093 CCACACACGAGCGGAGCAGGGTGCAGGACAGATGACCTGGTGAAGATGCTCTTCAGCCCA 2152  
701 GluGluMetAspLeuSerGlnGluGlnProLeuAspAlaGlnGlnGlyProProGluPro 720  
2153 GAAGAGATGGACCTCTCCAGGAGCGCCCTGGATGCCAGCAGGCGCCCCCGAACCT 2212  
721 AlaGlnGluSerLeuSerGlySerGluSerLysProLysAspGluLeu 736  
2213 GCACAAGAGTCTCTCTCAGGAGTGAATCGAAGCCCAAGGATGAGCTA 2260

## RESULT 8

AK027697 2585 bp mRNA linear PRI 01-AUG-2002  
LOCUS Homo sapiens cDNA FLJ14791 fis, clone NT2RP4001064, weakly similar  
DEFINITION to SYNAPTONEMAL COMPLEX PROTEIN SC65.

AK027697  
ACCESSION AK027697.1 GI:14042570

VERSION oligo capping; fis (full insert sequence).

KEYWORDS Homo sapiens (human)

SOURCE Homo sapiens

ORGANISM Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

## REFERENCE

## AUTHORS

1 Isogai,T., Ota,T., Hayashi,K., Sugiyama,T., Otsuki,T., Suzuki,Y.,  
Nishikawa,T., Nagai,K., Sugano,S., Aotsuka,S., Yoshikawa,Y.,  
Matsumura,H., Ishii,S., Kawai,Y., Saito,K., Yamamoto,J.,  
Wakamatsu,A., Nakamura,Y., Nagahara,K., Masuho,Y. and Sasaki,N.  
NEDO human cDNA sequencing project

Unpublished

2 (bases 1 to 2585)

Isogai,T. and Otsuki,T.

Direct Submission

Submitted (10-MAY-2001) Takao Isogai, Helix Research Institute,

Genomics Laboratory; 1532-3 Yana, Kisarazu, Chiba 292-0812, Japan

(E-mail:genomics@hri.co.jp, Tel:81-438-52-3975, Fax:81-438-52-3986)

NEDO human cDNA sequencing project supported by Ministry of

Economy, Trade and Industry of Japan; cDNA full insert sequencing,

Research Association for Biotechnology; cDNA library construction,

5'- & 3'-end one pass sequencing and clone selection; Helix

Research Institute (supported by Japan Key Technology Center etc.)

and Department of Virology, Institute of Medical Science,

University of Tokyo.

Location/Qualifiers

## FEATURES

## source

1..2585  
/organism="Homo sapiens"  
/mol\_type="mRNA"  
/db\_xref="taxon:9606"  
/clone="NT2RP4001064"  
/cell\_line="NT2"  
/cell\_type="teratocarcinoma"  
/clone\_lib="NT2RP4"  
/note="Cloning vector: pME18SFL3-mRNA from NT2 neuronal  
precursor cells after 2-weeks retinoic acid (RA)  
induction."  
53..2263  
/note="unnamed protein product"

## CDS

/codon\_start=1  
/protein\_id="BAB55305.1"  
/db\_xref="GI:14042571"

/translation="MAVRALKLTLLAVAAASQAESEVSEAGWMTVDLLFABGTA  
AVARGDPCWLVSMERALEPAAALRLRCCTCAADFPWELDDPDWSPSPAAGSAA  
ALRDLSPFGGLLRRAACLRCLGPPAAHSLSEMELEPKRSPYNLYQVAFPKINLE



KAVAAATFFVGNPEHMEQNLDDYYQTMSCVKEADPKDLQTPHMQEPRLGVRLYSE  
 EOPQAVPHLBAUQYEVAYEBCALCEGPDYDGNLYLEYNADLFOALTDHYIQL  
 NCKQCVTELASHSPREKFFEDFUPSHYNYLOFAYNIGNYQVACAKYLLFFPND  
 EVMQNLYAYAMLEETRSIGPESAKYRSLLEKELLFPAYDVGIFPVPDPS  
 WTPBEVPIKQEQKSERETAVRISQIGNLMEIETLVEEKTBSLDVSRLTREGG  
 PLYBEGISLTWNSKLNSORVMDGVTSDHECOLRLTNVAATSGDGYRGOTSPHT  
 PNEKPYGVTFKALKLQGGKVPLOQSAHLYNTEKVRIMESVRLDTPLYSYSHL  
 VCRFAIEVQAEKDDSHPVHVDACILNAETLVCVKEPPAYTRDYSAIILYNGDFPG  
 CNFPTEDAKITVAVQPCQRAVFGSSGTENPHGVKAVTRGORCAIALWFLDPRH  
 SERDRVQADDLLFKMLFSPREMDLSOEQPLDAQQGPPEFAQESUSGSSSKPKDEL"

## ORIGIN

## Alignment Scores:

Pred. No.: 1,896-293 Length: 2585  
 Score: 3846.00 Matches: 733  
 Percent Similarity: 99.59% Conservative: 0  
 Best Local Similarity: 99.59% Mismatches: 3  
 Query Match: 99.38% Indels: 0  
 DB: 9 Gaps: 0

US-10-045-815-4 (1-736) x AK027697 (1-2585)

QY	1	MetAlaValArgAlaLeuLysLeuLeuThrThrLeuLeuAlaValAlaAlaSer	20
DB	53	ATGGCGGTACCGCGCTTGAAGCTGCTGACCACTGCTGGCTGCTGGCGGCTGCCCTC	112
QY	21	GlnAlaGluValGluSerGluAlaGlyTrpGlyMetValThrProAspLeuLeuPheAla	40
DB	113	CAGCCGAGGTTCAGTCCGAGCAGGATGGGCGATGGTACGCCCTGATCTGCTTTCGCC	172
QY	41	GluGlyThrAlaAlaTyrAlaArgGlyAspTrpProGlyValValLeuSerMetGluArg	60
DB	173	GAGGGGACCGGACCTACCGCGCGGGGACTGGCCCGGGGTGGTCTCTGAGCATGGAACGG	232
QY	61	AlaLeuArgSerArgAlaAlaLeuArgAlaLeuArgLeuArgCysArgThrGlnCysAla	80
DB	233	CGCTGCGCTCCCGGCGAGCCCTCCGCGCCCTTCGCTGCGCTCCGCGACCCAGTGTGCC	292
QY	81	AlaAspPheProTrpGluLeuAspProAspTrpSerProSerProAlaGlnAlaSerGly	100
DB	293	GCCGACTTCCCGTGGAGCTGGACCCCGACTGGTCCCCCGCCGCGCCAGGCTCGGGC	352
QY	101	AlaGlyAlaLeuArgAspLeuSerPhePheGlyGlyLeuLeuArgArgAlaAlaCysLeu	120
DB	353	CGCCGCGCCCTCGCGGACCTGAGCTTCTTCGGGGCCCTTCGCGTCCGCTGCGCTGCGCTG	412
QY	121	ArgArgCysLeuGlyProProAlaAlaHisSerLeuSerGluGluMetGluLeuGluPhe	140
DB	413	CGCCGCTGCTCGGGCCGCGCGCCGCTGCTGCTGAGAGAGATGGAGTGGAGTTC	472
QY	141	ArgLysArgSerProTyrAsnTyrLeuGlnValAlaTyrPheLysIleAsnLysLeuGlu	160
DB	473	CGCAAGCGGAGCCCTCAACTACTCTGAGTCCGCTACTTCAAGATCAACAAAGTTGGAG	532
QY	161	LysAlaValAlaAlaAlaHisThrPhePheValGlyAsnProGluHisMetGluMetGln	180
DB	533	AAAGCTGTGTGAGACACACACCTTCTTGTGGGCAATCTTGAGCACATGGAAATGCAG	592
QY	181	GlnAsnLeuAspTyrTyrGlnThrMetSerGlyValLysGluAlaAspPheLysAspLeu	200
DB	593	CAGAACCTTAGACTATTCAAAACCATGCTCGAGTGAAGGAGGCGGACTTCAAGGATCTT	652
QY	201	GluThrGlnProHisMetGlnGluPheArgLeuGlyValArgLeuTyrSerGluGluGln	220
DB	653	GAGACTCAACCCCATATGCAAGAAATTCGACTGGAGTGGGACTCTTACTCAGAGGAACAG	712
QY	221	ProGlnGluAlaValProHisLeuGluAlaAlaLeuGlnGluTyrPheValAlaTyrGlu	240
DB	713	CCACAGGAAGCTGTGCCCCACCTAGAGGCGGCGCTGCAAGAAATCTTTGTGGCCTATGAG	772
QY	241	GluCysArgAlaLeuCysGluGlyProTyrAspTyrAspGlyTyrAsnTyrLeuGluTyr	260
DB	773	GAGTGGCTGCTCTCGGAAGGGCCCTATGACTACGATGGCTACACCTTGTAGTAC	832

QY	261	AsnAlaAspLeuPheGlnAlaIleThrAspHisTyrIleGlnValLeuAsnCysLysGln	280
DB	833	AACGCTGACCTTTCAGGCCCATCACAGATCATTATCATCCAGTCTCAACTCAAGTAAAGCAG	892
QY	281	AsnCysValThrGluLeuAlaSerHisProSerArgGluLysProPheGluAspPheLeu	300
DB	893	AACCTGTTCAGGAGCTTCTCCCAACCAAGTCAGAGAGCCCTTTGAAGACTTCTCTC	952
QY	301	ProSerHisTyrAsnTyrLeuGlnPheAlaTyrTyrAsnIleGlyAsnTyrThrGlnAla	320
DB	953	CCATCGCATATAATATCTGCAGTTTGCTACTATAACATTGGGAATTTATACAGGCT	1012
QY	321	GlyGluCysAlaLysThrTyrLeuLeuPhePheProAsnAspGluValMetAsnGlnAsn	340
DB	1013	GTTGAATGTGCCAAGACCTATCTCTCTCTTCCCAATGACGAGTGATGAACCAAAAT	1072
QY	341	LeuAlaTyrTyrAlaAlaMetLeuGlyGluGluHisThrArgSerIleGlyProArgGlu	360
DB	1073	TTGGCTATTATGACGCTATGCTTGGAGAAGAACACACAGATCCATCGGCCCGGTGAG	1132
QY	361	SerAlaLysGluTyrArgGlnArgSerLeuLeuLysGluLeuLeuPhePheAlaTyr	380
DB	1133	AGTGCCCAAGGAGTACCGACGAGGCTTACTTGAAAAAGAACTGCTTTCTTCGCTTAT	1192
QY	381	AspValPheGlyIleProPheValAspProAspSerTyrThrProGluGluValIlePro	400
DB	1193	GATGTTTTTGGAAATCCCTTTGTGGATCCGGATTCATGGACTCCAGAGAAGTGTATCC	1252
QY	401	LysArgLeuGlnGluLysGlnLysSerGluArgGluThrAlaValArgIleSerGlnGlu	420
DB	1253	AAGAGATTCAAGAGAGAAACAGAAAGTCAGAACCGGAAACAGCGTACGCGATCTCCAGGAG	1312
QY	421	IleGlyAsnLeuMetLysGluIleGluThrLeuValGluGluLysThrLysGluSerLeu	440
DB	1313	ATTGGGAACCTTATGAAGAAATCGAGACCCCTTGTGGAGAGAAACCAAGAGTCACTG	1372
QY	441	AspValSerArgLeuThrArgGluGlyProLeuLeuTyrGluGlyIleSerLeuThr	460
DB	1373	GATGTGAGCAGACTGACCCGGGAAGTGGCCCTGCTGTATGAAGGCATCAGTCTCACC	1432
QY	461	MetAsnSerLysLeuLeuAsnGlyTyrGlnArgValValMetAspGlyValIleSerAsp	480
DB	1433	ATGAATCAAACTCCTGAATGGTTTCCAGCGGGTGGTGGAGCGGCTTATCTCTGAC	1492
QY	481	HisGluCysGlnGluLeuGlnArgLeuThrAsnValAlaAlaThrSerGlyAspGlyTyr	500
DB	1493	CACGAGTGTGAGAGCTGACAGAGACTGACCAATGTGGCAGCAACCTCAGAGAGATGGCTAC	1552
QY	501	ArgGlyGlnThrSerProHisThrProAsnGluLysPheTyrGlyValThrValPheLys	520
DB	1553	CGGGGTGAGACTCCCCACATCTCCCAATGAAAGTTCTATGGTGTCACTGTCTTCAAA	1612
QY	521	AlaLeuLysLeuGlyGlnGluLysValProLeuGlnSerAlaHisLeuTyrTyrAsn	540
DB	1613	GCCTCAAGCTGGGCAAGAGGCAAGTTCTCTGACAGAGTGGCCCTGCTACTACAC	1672
QY	541	ValThrGluLysValArgIleMetGluSerTyrPheArgLeuAspThrProLeuTyr	560
DB	1673	GTGACGGAAGGTCGCGCGCATCATCTGAGTCTTCCGCTCGGATGATGATGATGATGATG	1732
QY	561	PheSerTyrSerHisLeuValCysArgThrAlaIleGluGluValGlnAlaGluArgLys	580
DB	1733	TTTTCTTCTCTCATCTGTGTGTGCGCACTGCCATCGAAGAGGTCGAGCAGAGAGGAG	1792
QY	581	AspAspSerHisProValHisValAspAsnCysIleLeuAsnAlaGluThrLeuValCys	600
DB	1793	GATGATAGTATCCAGTCCAGTCCAGTCCAGTCCAGTCCAGTCCAGTCCAGTCCAGTCCAG	1852
QY	601	ValLysGluProProAlaTyrThrPheArgAspTyrSerAlaIleLeuTyrLeuAsnGly	620
DB	1853	GTCAAAGAGCCCCCAGGCTACACCTTCCGCGACTACAGCGCCATCTTACTTAATGGG	1912



621 AspPheAspGlyGlyAsnProPheThrGluLeuAspAlaLysThrValThrAlaGlu 640  
1913 GACTTCGATGGCGGAACITTTATTTCTACTGAACTGGATGCCAAGACCGTGACGGCAGAG 1972  
641 ValGlnProGlnCysGlyArgAlaValGlyPheSerSerGlyThrGluAsnProHisGly 660  
1973 GTGACGCTCAGTGTGAGAGCGGTGGGATTTCTTTCAGGCACTGAAACCCACATGGA 2032  
661 ValIysAlaValThrArgGlyGlnArgCysAlaIleAlaLeuTrpPheThrLeuAspPro 680  
2033 GTGAAGGCTGTACACGAGGGGACGCTGTGCCATGGCCCTGTGGTTCACCTGGACCT 2092  
681 ArgHisSerGluArgAspArgValGlnAlaAspLeuValIysMetLeuPheSerPro 700  
2093 CGACACAGCGAGCGGACAGGGTGCAGGACAGTGCCTGGTGAAGATGCTCTTCAGCCCCA 2152  
701 GluLeuMetAspLeuSerGlnGlnProLeuAspAlaGlnGlnGlyProProGluPro 720  
2153 GAAGAGATGACCTCTCCAGGAGCAGCCCTGATGCCAGAGGGCCCCCCCCAACCT. 2212  
721 AlaGlnGluSerLeuSerGlySerGluSerIysProIysAspGluLeu 736  
2213 GCACAGAGTCTCTCTCAGCGCTGATCGAGCCCAAGCATGAGCTA 2260

RESULT 9  
AX877139 2524 bp DNA linear PAT 17-DEC-2003  
LOCUS Sequence 12044 from Patent EP1074617.  
DEFINITION AX877139  
ACCESSION AX877139  
VERSION AX877139.1 GI:40031875  
KEYWORDS  
SOURCE Homo sapiens (human)  
ORGANISM Homo sapiens  
Eukaryota; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
1  
Ota, T., Isogai, T., Nishikawa, T., Hayashi, K., Saito, K., Yamamoto, J.,  
Ishii, S., Sugiyama, T., Wakamatsu, A., Nagai, K. and Otsuki, T.  
Primers for synthesising full-length cDNA and their use  
Patent: EP 1074617-A 12044 07-FEB-2001;  
Research Association for Biotechnology (JP)  
FEATURES  
source  
1..2524  
/organism="Homo sapiens"  
/mol\_type="unassigned DNA"  
/db\_xref="taxon:9606"  
16..2226  
/note="unnamed protein product"  
/codon\_start=1  
/protein\_id="CAE89589.1"  
/db\_xref="GI:40031875"  
/translation="MAVRLKLLTLLAVVAASQAEVSEAGGMVTPDLLPAEGTA  
AYARGDWGVVLSRRAALRLRRCRTCAADFPWELDPNSPFAQASGA  
ALRDLSPFGGLRRAALRLRCLGPPAAHSSEMELEFRKRSPNYILOVAFYKINKLE  
KAVAAHTFFVGNPEHMQNLQNDYQTMSSGKEADFQLETPHMQEFLRGVLYSE  
EQPEAVPHEALQVEYVAYEECRALCEGPDYDGYNLEYNADLFQAITDHYIOVL  
NCKQNVTEALSHSREKPFDFLPVSHYVLOPANYIGNYTOAVECAKYLILFFND  
EVMQNLAAYAAMLGEEHTRSIGPESRKEVORSLLEKLLFFAVDVGIPFVDPDS  
WTPPEVTPRIQEKQSERATVARIQSIGNLMKEIETLVEEKTKEISLDYSLRTREGG  
PLLYEGISLTNKLNSQVMDGVISDHECEQLRLINVAISGDSGVRGQTSPT  
PNEKPYGVTFKALQGEKVPLOQSAHLYNVTKEVRNIMESYRLDTPLYSYSL  
VCRTAIEVQKDDSPHVDNCLNAETLVCKEPPAYTFRDYSAIYLNDGDFG  
GNFYETLDAKTVTAEOVQCGRAVFGSSGTENPHGVKAVTRGQRCALWFLDPRH  
SERDRVQADDLVKMLFSPEMDLSQEQPMDAQGPPEPAQESLSGSSESKPDEL"

Alignment Scores:  
Pred. No.: 2,63e-293 Length: 2524  
Score: 3844.00 Matches: 732  
Percent Similarity: 99.59% Conservative: 1  
Best Local Similarity: 99.46% Mismatches: 3  
Query Match: 99.33% Indels: 0  
DB: 6 Gaps: 0

US-10-045-815-4 (1-736) x AX877139 (1-2524)  
QY 1 MetAlaValArgAlaLeuLysLeuLeuThrThrLeuLeuAlaValValAlaAlaSer 20  
Db 16 ATGGCGGTACGGCGGTGAAGCTCTGACCACTGCTGGTGTCTGCTGGCGCTGCCTCC 75  
QY 21 GlnAlaGluValGluSerGluAlaGlyTrpGlyMetValThrProAspLeuLeuPheAla 40  
Db 76 CAAAGCCGAGGTGAGTCCGAGCAGGATGGGGCAATGGTGAAGCCCTGATCTGCTCTCTGCC 135  
QY 41 GluGlyThrAlaAlaValAlaArgGlyAspTrpProGlyValValLeuSerMetGluArg 60  
Db 136 GAGGGAGCGGACGCTACGCGCGGGGACTGGCCCGGGGTGGTCTCTGAGCATGAAACGG 195  
QY 61 AlaLeuArgSerArgAlaAlaLeuArgAlaLeuArgLeuArgCysArgThrGlnCysAla 80  
Db 196 GCGCTGCGCTCCGCGCAGCCCTCGCGCCCTTCGCCCTGCGCTGCGCAGCCAGTGTGCC 255  
QY 81 AlaAspPheProTrpGluLeuAspProAspTrpSerProSerProAlaGlnAlaSerGly 100  
Db 256 GCGGACTTCCCGTGGGAGCTGGACCCGACTGGTCCCGACGCCCGCCAGGCTCGGCG 315  
QY 101 AlaGlyAlaLeuArgAspLeuSerPheGlyGlyLeuLeuArgArgAlaAlaCysLeu 120  
Db 316 GCGCGCGCCCTCGCGCAGCCTGAGCTTCTCGGGGCGCTTCTGCGTGGGCTGCGCTGCC 375  
QY 121 ArgArgCysLeuGlyProProAlaAlaHisSerLeuSerGluMetGluLeuGluPhe 140  
Db 376 CGCGCTGCTCGGCGCGCGCGCCACTCGCTCAGGAAGAGATGAGAGTGGAGTTC 435  
QY 141 ArgLysArgSerProTrpAsnTrpLeuGlnValAlaValPheLysIleAsnLysLeuGlu 160  
Db 436 CGCAGCGGAGCGCCCTACCACTACCTGCGAGTTCGCTACTTCAAGATCAACAAGTTGGAG 495  
QY 161 LysAlaValAlaAlaAlaHisThrPhePheValGlyAsnProGluHisMetGluMetGln 180  
Db 496 AAAGCTGTTGCTGCGACACACACCTTCTTCGTGGGCAATCCTGAGCAGACATGAAATGCGAG 555  
QY 181 GlnAsnLeuAspTrpTrpGlnThrMetSerGlyValLysGluAlaAspPheLysAspLeu 200  
Db 556 CAGAACCTAGACTATTACCAACCAATGCTGAGTGAAGAGCGGCGACTTCAAGGATCTT 615  
QY 201 GluThrGlnProHisMetGlnGluPheArgLeuGlyValArgLeuTrpSerGluGluGln 220  
Db 616 GAGACTCAACCCCATATGCAAGAAATTCGACTGGGAGTCCGACTCTACTTCAGAGGAACAG 675  
QY 221 ProGlnGluAlaValProHisLeuGluAlaLeuGlnGluTrpPheValAlaValGlu 240  
Db 676 CCACAGGAAGCTGTGCCCCCAGCTAGAGGCGCGCTCCAAAGATATCTTTGTGGCTATGAG 735  
QY 241 GluCysArgAlaLeuCysGluGlyProTrpAspTrpAspGlyTrpAsnTrpLeuGluTrp 260  
Db 736 GAGTGGCGTGGCTCTCGAGAGGCGCCCTATGACTAGTGGCTTACAACTACTCTTGAGTAC 795  
QY 261 AsnAlaAspLeuPheGlnAlaIleThrAspHisTrpIleGlnValLeuAsnCysLysGln 280  
Db 796 AACGCTGACCTCTTCCAGGCCATCAGATCATATCATCAGGTCTCTCACTGTAAGCAG 855  
QY 281 AsnCysValThrGluLeuAlaSerHisProSerArgGluLysProPheGluAspPheLeu 300  
Db 856 AACTGTGTACGAGGCTGTGTTCCACCCAGTCCGAGAGAGCCCTTTGAGAGACTTCCCTC 915  
QY 301 ProSerHisTrpAsnTrpLeuGlnPheAlaTrpAsnIleGlyAsnTrpThrGlnAla 320  
Db 916 CCATGCGATTATAATTATCTGCAAGTTTGGCTACTATAACATTGGGAATTATACACAGGCT 975  
QY 321 GlyGluCysAlaLysThrTrpLeuLeuPhePheProAsnAspGluValMetAsnGlnAsn 340  
Db 976 GTTGAATGTGCCAAGACCTATCTTCTTCTTCTTCCCAATGACGAGGTGATGAACCAAAAT 1035  
QY 341 LeuAlaValTrpAlaAlaMetLeuGluGlnHisThrArgSerIleGlyProArgGlu 360

DB 1036 TTGGCCCTATTATGACAGTATGCTTGGAGAGAACACACAGATCCATCGGCCCGCGTGAG 1095  
QY 361 SerAlaLysGluTyrArgGlnArgSerLeuLeuGluLysGluLeuPhePheAlaTyr 380  
DB 1096 AGTCCCAAGGAGTACCGACAGCGAAGCTTACTGGAAAGAAAGTCTTTCCTCGTTAT 1155  
QY 381 AspValPheGlyIleProPheValAspProAspSerTyrThrProGluGluValIlePro 400  
DB 1156 GATGTTTGGAAATCCCTTTGGATCCGGATTATGGAATCCAGAGTCCAGAAAGTATCC 1215  
QY 401 LysArgLeuGlnGluLysGlnLysSerGluArgGluThrAlaValArgLysSerGlnGlu 420  
DB 1216 AAGAGATTCCAGAGAGAAACAGAGTCCAGACCGGAAACAGCCGTACGATCTCCAGGAG 1275  
QY 421 IleGlyAsnLeuMetLysGluIleGluThrLeuValGluGluLysThrLysGluSerLeu 440  
DB 1276 ATTGGGAACCTTATGAAGGAATTCGAGACCTTGTGGAAGAGAGAACCAAGGAGTCACTG 1335  
QY 441 AspValSerArgLeuThrArgGluGlyProLeuLeuTyrGluGlyIleSerLeuThr 460  
DB 1336 GATGTGAGCAGACTGACCCGGAAGGTGGCCCTCTGTATGAAGGATCAGTCTCACC 1395  
QY 461 MetAsnSerLysLeuLeuAsnGlyTyrGlnArgValValMetAspGlyValIleSerAsp 480  
DB 1396 ATGAACCTCCAAACTCCTGAATGTTCCAGCGGTGTGTGATGGACGGGTAACTCTCTGAC 1455  
QY 481 HisGluCysGlnGluLeuGlnArgLeuThrAsnValAlaAlaThrSerGlyAspGlyTyr 500  
DB 1456 CACGAGTGCAGAGCTGCAGAGATGACCAATGTGGCAGCAACCTTCAGGAGATGGCTAC 1515  
QY 501 ArgGlyGlnThrSerProHisThrProAsnGluLysPheTyrGlyValThrValPheLys 520  
DB 1516 CGGGGTACAGACCTCCCACTATCTCCCAATGAAGTCTATGTGTCTACGTCTTCAA 1575  
QY 521 AlaLeuLysLeuGlyGlnGluLysValProLeuGlnSerAlaHisLeuTyrTyrAsn 540  
DB 1576 GCGCTCAAGCTGGGCAAGAGGCAAAAGTCTCTGCAGAGTGCACCTGCTACTACAAC 1635  
QY 541 ValThrGluLysValArgGlyMetGluSerTyrPheArgLeuAspThrProLeuTyr 560  
DB 1636 GTGACGGAGAGGTCCGGCGCATATGAGTCTTCTTCCCGCTGGATGACCCCTCTAC 1695  
QY 561 PheSerTyrSerHisLeuValCysArgThrAlaIleGluGluValGlnAlaGluArgLys 580  
DB 1696 TTTTCTACTCTCATCTGTGTGGCGCACTGCCATCGAAGAGTCCAGGCGAGAGGAAG 1755  
QY 581 AspAspSerHisProValHisValAspAsnCysIleLeuAsnAlaGluThrLeuValCys 600  
DB 1756 GATGATAGTATCAGTCCAGTCCAGTGACACTGCACTCTGTAATCCGAGACCCCTCGTGT 1815  
QY 601 ValLysGluProProAlaTyrThrPheArgAspTyrSerAlaIleLeuTyrLeuAsnGly 620  
DB 1816 GTCAAAGAGCCCGACGCTACCTTCGCGACTACAGCGCCATCTTTTACCTAAATGGG 1875  
QY 621 AspPheAspGlyGlyAsnPheTyrPheThrGluLeuAspAlaLysThrValThrAlaGlu 640  
DB 1876 GACTTCGATGGCGGAACCTTTATTTTCACTGAATCGATCCAGACCTGAAAACCCACATGGA 1935  
QY 641 ValGlnProGlnCysGlyArgAlaValGlyPheSerSerGlyThrGluAsnProHisGly 660  
DB 1936 GTGCAGCCTCAGTGTGGAAGAGCGGTGGATCTCTTCAGGCATCGAATCCACATGGA 1995  
QY 661 ValLysAlaValThrArgGlyGlnArgCysAlaIleAlaLeuThrPheThrLeuAspPro 680  
DB 1996 GTGAAGGCTGTACACAGGGGCGACGCTGTGCCATCGCTGTGGTTCACCTGGACCT 2055  
QY 681 ArgHisSerGluArgAspArgValGlnAlaAspAspLeuValLysMetLeuPheSerPro 700  
DB 2056 CGACACAGCGAGCGGACAGGTCAGCAGATGACTGTGTGAAGATGCTCTTCAGCCCA 2115  
QY 701 GluGluMetAspLeuSerGlnGluGlnProLeuAspAlaGlnGlnGlyProGluPro 720  
DB 2116 GAAGAGATGGACCTCTCCAGGAGCAGCCCATGATGCCAGGAGGCGGCCCGGAACT 2175

QY 721 AlaGlnGluSerLeuSerGlySerGluSerLysProLysAspGluLeu 736  
DB 2176 GCACAGAGTCTCTCTCAGGCACTGAATCGAAGCCCAAGGATGAGCTA 2223  
RESULT 10  
BD156505 2524 bp DNA linear PAT 17-JAN-2003  
LOCUS Primer for synthesizing full-length cDNA and use thereof.  
DEFINITION  
ACCESSION BD156505  
VERSION BD156505.1 GI-27862263  
KEYWORDS JP 2002191363-A/11348.  
SOURCE Homo sapiens (human)  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
REFERENCE 1 (bases 1 to 2524)  
AUTHORS Ota,T., Isogai,T., Nishikawa,T., Hayashi,K., Saito,K., Yamamoto,J.,  
Ishii,S., Sugiyama,T., Wakamatsu,A., Nagai,K. and Otsuki,T.  
TITLE Primer for synthesizing full-length cDNA and use thereof  
JOURNAL Patent: JP 2002191363-A 11348 09-JUL-2002;  
HELIX RESEARCH INSTITUTE  
COMMENT OS Homo sapiens (human)  
PN JP 2002191363-A/11348  
PD 09-JUL-2002  
PF 28-JUL-2000 JP 2000280990  
PI TOSHIO OTA, TAKAO ISOGAI, TETSUO NISHIKAWA, KOJI HAYASHI, KAORU  
PI JUNICHI YAMAMOTO, SHIZUKO ISHII, TOMOYASU SUGIYAMA, AI WAKAMATSU,  
PI KEIICHI NAGAI, TETSUJI OTSUKI  
PC C12N15/09, C07K14/47, C07K16/18, C12N1/15, C12N1/19, C12N1/21, C12N5/ PC  
10, C12P21/02, C12Q1/69//C12P21/08, G06F17/30, C12N15/00, C12N5/00 CC  
Primer for synthesizing full-length cDNA and use thereof FH Key  
Location/Qualifiers (16) .. (2223).  
FT CDS Location/Qualifiers  
source 1..2524  
/organism="Homo sapiens"  
/mol\_type="genomic DNA"  
/db\_xref="taxon:9606"  
ORIGIN  
Alignment Scores:  
Pred. No.: 2,63e-293 Length: 2524  
Score: 3844.00 Matches: 732  
Percent Similarity: 99.59% Conservative: 1  
Best Local Similarity: 99.46% Mismatches: 3  
Query Match: 99.33% Indels: 0  
DB: 6 Gaps: 0  
US-10-045-815-4 (1-736) x BD156505 (1-2524)  
QY 1 MetAlaValArgAlaLeuLysLeuLeuThrThrLeuLeuAlaValAlaAlaSer 20  
DB 16 ATGGCGGTACGCGCTTGAAGCTGTGACACACTGCTGGCTGTCTGGCGCGCTGCCCTCC 75  
QY 21 GlnAlaGluValGluSerGluAlaGlyTyrGlyMetValThrProAspLeuPheAla 40  
DB 76 CAAAGCCGAGGTGAGTCCGAGCAGGATGGGCGATGGGACCTGATCTCTTTCGCC 135  
QY 41 GluGlyThrAlaAlaTyrAlaArgGlyAspTyrProGlyValValLeuSerMetGluArg 60  
DB 136 GAGGAGACCGCAGCCTACGCGCGCGGACTGGCCCGGGTGGCTCTGAGCATGGAACGG 195  
QY 61 AlaLeuArgSerArgAlaAlaLeuArgAlaLeuArgLeuArgCysArgThrGlnCysAla 80  
DB 196 GCGCTGCGCTCCCGGAGAGCCCTCCGCGCCCTTCGCTGGCTCGCGACCATGATGTGCC 255  
QY 81 AlaAspPheProTyrGluLeuAspProAspTyrSerProAlaGlnAlaSerGly 100  
DB 256 GCCGACTTCCGTTGGAGTGGACCGCCCGGACTGTGTCCTCCCGAGCCCGCCGCTCGGCC 315

101 AlaGlyAlaLeuArgAspLeuSerPhePheGlyGlyLeuLeuArgAlaAlaCysLeu 120  
102 |||||  
316 GCCGCCCTCGCGACCTCTCTCGGGGCTTCTCGGTCGCGCTCGCTCGCTCG 375  
121 ArgArgCysLeuGlyProProAlaAlaHisSerLeuSerGluGluMetGluLeuGluPhe 140  
122 |||||  
376 CGCGCTGCTCGGCCCGCGCGGCCCACTCGCTCAGCGAAGAGATGGAGTTC 435  
141 ArgLysArgSerProTyrAsnTyrLeuGlnValAlaTyrPheLysLysLeuLeuGlu 160  
142 |||||  
436 CGCAAGCGGAGCCCTCAACTACTCTGAGTGGCTCTCAAGATCAACAAGTGGAG 495  
161 LysAlaValAlaAlaHisThrPhePheValGlyAsnProGluHisMetGluMetGln 180  
162 |||||  
496 AAGCTGTGCTGCGACACACACCTCTCTCGTGGCAATCTCTGAGCACATGGAAATGCAG 555  
181 GlnAsnLeuAspTyrTyrGlnThrMetSerGlyValLysGluAlaAspPheLysAspLeu 200  
182 |||||  
556 CAGAACCTAGACTATTACCAACCATGTCTGGAGTGAAGAGCGGCTTCAAGGATCTT 615  
201 GluThrGlnProHisMetGlnGluPheArgGlyValArgLeuTyrSerGluGluGln 220  
202 |||||  
616 GAGACTCAACCCCATATGCAAGAAATTCGACTGGAGTGGACTCTACTCAGAGGAACAG 675  
221 ProGlnGluAlaValProHisLeuGluAlaAlaLeuGlnGluTyrPheValAlaTyrGlu 240  
222 |||||  
676 CCACAGGAGCTGTGCCCACTAGAGCGCGCTGCAAGAAATCTTTGTGGCTATGAG 735  
241 GluCysArgAlaLeuCysGluGlyProTyrAspTyrAspGlyTyrAsnTyrLeuGluTyr 260  
242 |||||  
736 GAGTGGCTGCTCTGCGAAGGCGCTTATGACTACGATGCTACAACTACTCTTGAAGTCT 795  
261 AsnAlaAspLeuPheGlnAlaThrAspHisTyrLeGlnValLeuAsnCysLysGln 280  
262 |||||  
796 AACCTGACCTCTCCAGGCCATCACAGATCATTACATCCAGGTCTCTCAACTGTAAGCAG 855  
281 AsnCysValThrGluLeuAlaSerHisProSerArgGluLysProPheGluAspPheLeu 300  
282 |||||  
856 AACTGTGTCAGAGCTTGTCTCCACCAAGTCGAGAGAGCCCTTTGAAAGCTCTCTC 915  
301 ProSerHisTyrAsnTyrLeuGlnPheAlaTyrTyrAsnIleGlyAsnTyrThrGlnAla 320  
302 |||||  
916 CCATCGCATTAATATCTGAGTTTGCCTACTATACATGGAATATACACAGCT 975  
321 GlyGluCysAlaLysThrTyrLeuLeuPhePheProAsnAspGluValMetAsnGlnAsn 340  
322 |||||  
976 GTTGAATGTGCCAAGACCTATCTTCTCTTCCCAATGACGAGGTGATGAACCAAAAT 1035  
341 LeuAlaTyrTyrAlaAlaMetLeuGlyGluGluHisThrArgSerIleGlyProArgGlu 360  
361 |||||  
1036 TTGGCTATTATGCGACTATGCTTGAGAGAAACACACAGATCCATCGGCCCGGTGAG 1095  
361 SerAlaLysGluTyrArgGlnArgSerLeuLeuGluLysGluLeuPhePheAlaTyr 380  
381 |||||  
1096 AGTCCAGAGGTACCGACAGGAGCTTCTGGAAGAAAGAACTCTTTCTTGTGCTTAT 1155  
381 AspValPheGlyIleProPheValAspProAspSerTyrThrProGluGluValIlePro 400  
401 |||||  
1156 GATGTTTTTGAATTCCTTTTGTGATCGGATTCATGAGCTCCAGAGAAAGTGTATCCC 1215  
401 LysArgLeuGlnGluLysGlnLysSerGluArgGluThrAlaValArgIleSerGlnGlu 420  
421 |||||  
1216 AAGAGATTGCAGAGAAACAGAAAGTCAGAACCGGAAACAGCGCTACGCACTCTCCAGGAG 1275  
421 IleGlyAsnLeuMetLysGluIleGluThrLeuValGluGluLysThrLysGluSerLeu 440  
441 |||||  
1276 ATTGGGACCTTATGAAGAAATCGAGACCTTGTGGAAGAGAGAACCAAGAGGATCACTG 1335  
441 AspValSerArgLeuThrArgGluGlyGlyProLeuLeuTyrGluGlyLysSerLeuThr 460  
461 |||||  
1336 GATGTGAGCAGACTGACCCGGGAAGGTGGCCCCCTCTGTATGAGAGGATCACTCTCACC 1395

461 MetAsnSerLysLeuLeuAsnGlyTyrGlnArgValValMetAspGlyValIleSerAsp 480  
462 |||||  
1396 ATGAACCTCAAACTCTGAATGGTTCCAGAGGGGTGGTGAATGGACGGGTATCTCTCACC 1455  
481 HisGluCysGlnGluLeuArgLeuThrAsnValAlaAlaThrSerGlyAspGlyTyr 500  
482 |||||  
1456 CACGAGTGTACAGAGCTGCAGAGACTGACCAATGTGGCAGACACCTCAGGAGATGGCTAC 1515  
501 ArgGlyGlnThrSerProHisThrProAsnGluLysPheTyrGlyValThrValPheLys 520  
502 |||||  
1516 CGGGGTGACAGCTCTCCCACTACTCCCAATGAAAGTTCTATGGTGTCTACTGTCTTCAA 1575  
521 AlaLeuLeuLeuGlyGlnGluLysValProLeuGlnSerAlaHisLeuTyrTyrAsn 540  
522 |||||  
1576 GGCCTCAAGCTCGGGCAAGAGCAAGTTCTCTGACAGATGCCACCTGTACTACTACAC 1635  
541 ValThrGluLysValArgArgIleMetGluSerTyrPheArgLeuAspThrProLeuTyr 560  
561 |||||  
1636 GTGACGGAAGGTGGCGGCATCATGGAGTCTTCTTCCGCTGGATACGCCCTCTCTAC 1695  
561 PheSerTyrSerHisLeuValCysArgThrAlaIleGluGluValIleAlaGluArgLys 580  
581 |||||  
1696 TTTTCTACTCTCATCTGCTGTGCGCACTGCGCATCGAAGAGGTCCAGGACAGAGGAAG 1755  
581 AspAspSerHisProValHisValAspAsnCysIleLeuAsnAlaGluThrLeuValCys 600  
601 |||||  
1756 GATGATGATCATCCAGTCCAGTGGACAACTGCATCTCTGAATGCCAGACCTCTCGTGTCT 1815  
601 ValLysGluProProAlaTyrThrPheArgAspTyrSerAlaIleLeuTyrLeuAsnGly 620  
621 |||||  
1816 GTCAAGAGAGCCCGACCTACACCTTCCGCACTACAGCGCATCTTTTACCTAAATGGG 1875  
621 AspPheAspGlyGlyAsnPheTyrPheThrGluLeuAspAlaLysThrValThrAlaGlu 640  
641 |||||  
1876 GACTTCGATGGCGGAACCTTTTATTTCACTGAACTGATGCCAAGACCGGTGACGGCAGAG 1935  
641 ValGlnProGlnCysGlyArgAlaValGlyPheSerSerGlyThrGluAsnProHisGly 660  
661 |||||  
1936 GTGACGCTCAGTGTGGAAGAGCGTGGGATTTCTTCCAGCACTGAAACCCACATGGA 1995  
661 ValLysAlaValThrArgGlyGlnArgCysAlaIleAlaLeuThrPheThrLeuAspPro 680  
681 |||||  
1996 GTGAAGCTGTCAACAGGGGCGCGTGTGCCATCGCTCGCTGTGGTTACCTTGGACCT 2055  
681 ArgHisSerGluArgAspArgValGlnAlaAspAspLeuValLysMetLeuPheSerPro 700  
701 |||||  
2056 CGACACAGCGAGCGGACAGGGTGCAGGACAGATGACCTGTGTAAGATGCTTCTTCAGCCCA 2115  
701 GluGluMetAspLeuSerGlnGluGlnProLeuAspAlaGlnGlnGlyProProGluPro 720  
721 |||||  
2116 GAAGAGATGGACCTCTCCAGGAGCAGCCCATGGATGCCAGAGGGCCCCCGGAACT 2175  
721 AlaGlnGluSerLeuSerGlySerGluSerLysProLysAspGluLeu 736  
736 |||||  
2176 GCACAGAGTCTCTCTCAGGACGTGAATCGAAGCCCAAGGATGAGCTA 2223

RESULT 11  
AK027680 2524 bp mRNA linear PRI 01-AUG-2002  
LOCUS Homo sapiens cDNA FLJ14774 fis, clone NT2RP4000051, weakly similar  
DEFINITION to SYNAPTONEMAL COMPLEX PROTEIN SC65.  
ACCESSION AK027680  
VERSION AK027680.1 GI:14042539  
KEYWORDS oligo capping; fis (full insert sequence).  
SOURCE Homo sapiens (human)  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
1  
REFERENCE  
AUTHORS  
Isegai,T., Ota,T., Hayashi,K., Sugiyama,T., Otsuki,T., Suzuki,Y.,  
Nishikawa,T., Nagai,K., Sugano,S., Aotsuka,S., Yoshikawa,Y.,  
Matsunawa,H., Ishii,S., Kawai,Y., Saito,K., Yamamoto,J.,  
Wakamatsu,A., Nakamura,Y., Nagahara,K., Masuho,Y. and Sasaki,N.

NEDO human cDNA sequencing project  
 Unpublished  
 2 (bases 1 to 2524)  
 Isogai, T. and Otsuki, T.  
 Direct Submission  
 Submitted (10-MAY-2001) Takao Isogai, Helix Research Institute,  
 Genomics Laboratory, 1530-3 Yana, Kisarazu, Chiba 292-0812, Japan  
 (E-mail: genomics@hri.co.jp, Tel: 81-438-52-3975, Fax: 81-438-52-3986)  
 NEDO human cDNA sequencing project supported by Ministry of  
 Economy, Trade and Industry of Japan; cDNA full insert sequencing:  
 Research Association for Biotechnology; cDNA library construction,  
 5'- & 3'-end one pass sequencing and clone selection; Helix  
 Research Institute (supported by Japan Key Technology Center etc.)  
 and Department of Virology, Institute of Medical Science,  
 University of Tokyo  
 Location/Qualifiers  
 1..2524  
 /organism="Homo sapiens"  
 /mol\_type="mRNA"  
 /db\_xref="taxon:9606"  
 /clone="NT2RP4.000051"  
 /cell\_line="NT2"  
 /cell\_type="teratocarcinoma"  
 /clone\_lib="NT2RP4"  
 /note="cloning vector: pME18SFL3-mRNA from NT2 neuronal  
 precursor cells after 2-weeks retinoic acid (RA)  
 induction."  
 16..2226  
 /note="unnamed protein product"  
 /codon\_start=1  
 /protein\_id="BAB55291.1"  
 /db\_xref="GI:14042540"  
 /translation="MAVRAKLLTLLAVAAASQAESEAGHGMVTPDLLPAEGTA  
 AYARDWPGVLSMERALRLRRCRTCAADFPELDPDPSRPAQASGA  
 ALRDLSPFGGLLRRAALRCICGPPAAHLSSELMELFERKSPNYLQVAFKINKLE  
 KAVAAAH"FFVGNPEHMEMQNLDYYQTMSCVKEADFQLETPHMQEFLGLVRLYSE  
 EQQAEVHLSEALQEIYFAEECALCEGYDYGNYLLEYNALDFQAITDHYIQVL  
 NCKQNVTLASHSPKFPEDFLPSHNYLQFAYNIGNYQAVECAKTYLLFFPND  
 EVMQNLYAAMLEGEHTSRIIGPRAKEYRQSLLEKLLFFAYDVGIPFVDPDS  
 WTPREVIKRLQEKQRETAIVRSQIEIGNLKMEIETLVEEKTKESLDVRLTREGG  
 PLLVEGISTLWKLNGSORVMDGVI SDHECCOELQRLNVAATSDGVGGTQTSHT  
 PNEKFGYTVFALKLGGQKVPQSAHLVNTKRYRIMESVFLDITPLYSYSHL  
 VCRALAEVQAEKDDSHPVHVDNCILNAETLVCVKFPATFNDYSAILYNGDDFG  
 CNFTFELDAKTIVTAEPQCGRAVGSFSSNTENPHGVKAVTRGQRCAIALWFTLDPRH  
 SERDRVQADDLLVKMLFSPPEMDLSQEQPMDAQGPPEPAQESLSSESKPDEL"

CDS  
 16..2226  
 /note="unnamed protein product"  
 /codon\_start=1  
 /protein\_id="BAB55291.1"  
 /db\_xref="GI:14042540"  
 /translation="MAVRAKLLTLLAVAAASQAESEAGHGMVTPDLLPAEGTA  
 AYARDWPGVLSMERALRLRRCRTCAADFPELDPDPSRPAQASGA  
 ALRDLSPFGGLLRRAALRCICGPPAAHLSSELMELFERKSPNYLQVAFKINKLE  
 KAVAAAH"FFVGNPEHMEMQNLDYYQTMSCVKEADFQLETPHMQEFLGLVRLYSE  
 EQQAEVHLSEALQEIYFAEECALCEGYDYGNYLLEYNALDFQAITDHYIQVL  
 NCKQNVTLASHSPKFPEDFLPSHNYLQFAYNIGNYQAVECAKTYLLFFPND  
 EVMQNLYAAMLEGEHTSRIIGPRAKEYRQSLLEKLLFFAYDVGIPFVDPDS  
 WTPREVIKRLQEKQRETAIVRSQIEIGNLKMEIETLVEEKTKESLDVRLTREGG  
 PLLVEGISTLWKLNGSORVMDGVI SDHECCOELQRLNVAATSDGVGGTQTSHT  
 PNEKFGYTVFALKLGGQKVPQSAHLVNTKRYRIMESVFLDITPLYSYSHL  
 VCRALAEVQAEKDDSHPVHVDNCILNAETLVCVKFPATFNDYSAILYNGDDFG  
 CNFTFELDAKTIVTAEPQCGRAVGSFSSNTENPHGVKAVTRGQRCAIALWFTLDPRH  
 SERDRVQADDLLVKMLFSPPEMDLSQEQPMDAQGPPEPAQESLSSESKPDEL"

ORIGIN  
 Alignment Scores:  
 Pred. No.: 2,63e-293 Length: 2524  
 Score: 3844.00 Matches: 732  
 Percent Similarity: 99.59% Conservative: 1  
 Best Local Similarity: 99.46% Mismatches: 3  
 Query Match: 99.33% Indels: 0  
 DB: 9 Gaps: 0  
 US-10-045-815-4 (1-736) x AK027680 (1-2524)

QY 1 MetAlaValArgAlaLeuLysLeuLeuThrThrLeuLeuAlaValAlaAlaSer 20  
 Db 16 ATGGCGGTACGGGGTGAAGAGCTGCTACACACTGCTGGCTGCGCGCGCTGCTCC 75  
 QY 21 GlnAlaGluValGluSerGluAlaGlyTyrGlyMetValThrProAspLeuPheAla 40  
 Db 76 CAACCCGAGGTGCGATCGGAGCAGGATGGGATGTTGACGCTGATTCCTTCGCC 135  
 QY 41 GluGlyThrAlaAlaTyrAlaArgGlyAspTyrProGlyValValLeuSerMetGluArg 60  
 Db 136 GAGGGGACCCGAGCTACGCGCGCGGAGCTGCGCGCGGCTGCTGAGCATGGAACGG 195  
 QY 61 AlaLeuArgSerArgAlaAlaLeuArgAlaLeuArgGlyAspThrGlnCysAla 80  
 Db 196 GCGCTGCGCTCCCGGCGAGCCCTTCGCGCCCTTCGCGCTGCGCTGCGCGAGTGTGCC 255

QY 81 AlaAspPheProTyrGluLeuAspProAspTyrSerProSerProAlaGlnAlaSerGly 100  
 Db 256 GCCGACTTCGCGTGGAGCTGGACCCCGACTGGTCCCCCAGCCCGGCGCGCTCGGGC 315  
 QY 101 AlaGlyAlaLeuArgAspLeuSerPhePheGlyGlyLeuLeuArgAlaAlaCysLeu 120  
 Db 316 GCCCGCCCGCTGCGCGACCTGAGCTTCTTCGGGGCGCTTCTGCGTGGCTGCTGCTG 375  
 QY 121 ArgArgCysLeuGlyProProAlaAlaHisSerLeuSerGluGluMetGluLeuGluPhe 140  
 Db 376 CGCGCTGCTGCG 435  
 QY 141 ArgLysArgSerProTyrAsnTyrLeuGlnValAlaTyrPheLysLeuAsnLysLeuGlu 160  
 Db 436 CGCAAGCGGAGCGCGCTACAACTACCTGCGAGTGGCTTCTTCAAGATCAACAAGTTGGAG 495  
 QY 161 LysAlaValAlaAlaAlaHisThrPheValGlyValAsnProGluHisMetGluMetGln 180  
 Db 496 AAGCTGTTGCTGCGAGCAGCACACTTCTTCTGGGGCAATCTTGAGCAGATGGAATGCGAG 555  
 QY 181 GlnAsnLeuAspTyrTyrGlnThrMetSerGlyValLysGluAlaAspPheLysAspLeu 200  
 Db 556 CAGAACCTAGACTATTACCAAAACCATGTCTGGAGTGAAGAGGCGCGACTTCAAGGATCTT 615  
 QY 201 GluThrGlnProHisMetGlnGluPheArgLeuGlyValArgLeuTyrSerGluGluGln 220  
 Db 616 GAGACTCAACCCCATATGCAAGAAATTCGACTGGGAGTGGGACTTCTTACTCAGAGGAACAG 675  
 QY 221 ProGlnGluAlaValProHisLeuGluAlaAlaLeuGlnGlnLysPheValAlaTyrGlu 240  
 Db 676 CCACAGAAAGCTGTGCCCCCAGCTAGAGCGCGCTGCAAGATATCTTGTGCGCTATGAG 735  
 QY 241 GluCysArgAlaLeuCysGluGlyProTyrAspTyrAspGlyTyrAsnTyrLeuGluTyr 260  
 Db 736 GAGTGGCGTGGCTCTGCGAAGGCGCTATGACTAGTGGTCAACACTACCTTCAAGTAC 795  
 QY 261 AsnAlaAspLeuPheGlnAlaThrAspHisTyrIleGlnValLeuAsnCysLysGln 280  
 Db 796 AAGCTGACTCTCTCCAGGCCATCACAGATCATTACATCCAGTCTCTCACTGTGAAGAG 855  
 QY 281 AsnCysValThrGluLeuAlaSerHisProSerArgGluLysProPheGluAspPheLeu 300  
 Db 856 AACTGTGTACCGAGCTTGTCTCCACCCAGTCGAGAGAGAGCCCTTTCAGACATTCCTC 915  
 QY 301 ProSerHisTyrAsnTyrLeuGlnPheAlaTyrTyrAsnIleGlyAsnTyrThrGlnAla 320  
 Db 916 CCATCGCATTAATATCTGAGTTCGCTACTATACTATGGAATATATACAGGCT 975  
 QY 321 GlyGluCysAlaLysThrTyrLeuLeuPhePheProAsnAspGluValMetAsnGlnAsn 340  
 Db 976 GTTGAATGTGCAACACCTATCTTCTCTTCCCAATGACAGGTGATGAACCAAAAT 1035  
 QY 341 LeuAlaTyrTyrAlaAlaMetLeuGlyGluHisThrArgSerIleGlyProArgGlu 360  
 Db 1036 TTGGCTTATATGAGCTATGCTTGGAGAGAACACACCATCATCGGCCCCCGGTGAG 1095  
 QY 361 SerAlaLysGluTyrArgGlnArgSerLeuLeuLysGluLysGluLeuPhePheAlaTyr 380  
 Db 1096 AGTCCAAAGGAGTACCGACAGCAAGCTACTTGAAAGAAAGAACTGCTTCTTCTGCTTAT 1155  
 QY 381 AspValPheGlyIleProPheValAspProAspSerThrThrProGluGluValIlePro 400  
 Db 1156 GATGTTTTTGGAAATCCCTTTTGTGATCCCGATTCATGCACTCCAGAGAGAGTATTCCTC 1215  
 QY 401 LysArgLeuGlnGluLysGlnLysSerGluArgGluThrAlaValArgIleSerGlnGlu 420  
 Db 1216 AAGAGATTGCAAGAGAAACAGAGTCAAGACGGAACACCGCTACCATCTCCAGGAG 1275  
 QY 421 IleGlyAsnLeuMetLysGluIleGluThrLeuValGluGluLysThrLysGluSerLeu 440  
 Db 1276 ATTGGGAACTTATGAAGGAAATCGAGACCTTGTGGAAGAGAGAACCAAGAGTCACTG 1335  
 QY 441 AspValSerArgLeuThrArgGluGlyProLeuLeuTyrGluGlyIleSerLeuThr 460

Db 1336 GATGTGACGACATGACCGGAGGTGGCCCTGCTGATGAAGCATCATCTCCACC 1395  
Qy 461 MetAsnSerLysLeuLeuAsnGlyTyrGlnArgValValMetAspGlyValIleSerAsp 480  
Db 1396 ATGAACCTCCAACTCTGATGTTCCACGGGTGTGATGACGGCGTAATCTCTGAC 1455  
Qy 481 HisGluCysGlnGluLeuGlnArgLeuThrAsnValAlaIleThrSerGlyAspGlyTyr 500  
Db 1456 CACGAGTGTCTGAGGAGTGTGACAGACTGACCAATGTGCAGCAACCTCAGGAGATGGCTAC 1515  
Qy 501 ArgGlyGlnThrSerProHisThrProAsnGluLysPheTyrGlyValThrValPheLys 520  
Db 1516 CGGGGTGAGACTCCCCACATCTCCCAATGAAAGTTCTATGGTGTCACTGTCTTCAA 1575  
Qy 521 AlaLeuLysLeuGlyGlnGluLysValProLeuGlnSerAlaHisLeuTyrTyrAsn 540  
Db 1576 GCCCTCAAGCTGGGGCAAGGCAAGATTCCTCTGCAGAGTGCACCTGTACTACAAC 1635  
Qy 541 ValThrGluLysValArgArgIleMetGluSerTyrPheArgLeuAspThrProLeuTyr 560  
Db 1636 GTGACGGAGAAGGTGCGCGCATCATGAGTCTTACTTCCGCTGGATACGCCCTCTAC 1695  
Qy 561 PheSerTyrSerHisLeuValCysArgThrAlaIleGluGluValGlnAlaGluArgLys 580  
Db 1696 TTTTCCTACTCTCATCTGTGTGCGCACTGCCATCGAAGAGTCCAGGCAGAGGAAG 1755  
Qy 581 AspAspSerHisProValHisValAspAsnCysIleLeuAsnAlaGluThrLeuValCys 600  
Db 1756 GATGATAGTATCCAGTCCAGTCCAGCAACTGCATCTCTGATGCCGAGACCTCGTGTGT 1815  
Qy 601 ValLysGluProProAlaTyrThrPheArgAspTyrSerAlaIleLeuTyrLeuAsnGly 620  
Db 1816 GTCAAG 1875  
Qy 621 AspPheAspGlyGlyAsnPheTyrPheThrGluLeuAspAlaLysThrValThrAlaGlu 640  
Db 1876 GACTTCGATGGCGAAACTTTTATTTTCACTGAACTGATGATGCAAGACCGTGACGCAAG 1935  
Qy 641 ValGlnProGlnCysGlyArgAlaValGlyPheSerSerGlyThrGluAsnProHisGly 660  
Db 1936 GTGACAGCTCATGTGGAAGAGCGGTGGGATTCCTTCAGGACCTGAAACCCACATGGA 1995  
Qy 661 ValLysAlaValThrArgGlyGlnArgCysAlaIleAlaLeuTyrPheThrLeuAspPro 680  
Db 1996 GTGAAGGTGTGACAGAGGGGAGCGGTGCGCATCGCCCTGTGTTTCACTGAGACCT 2055  
Qy 681 ArgHisSerGluArgAspArgValGlnAlaAspAspLeuValLysMetLeuPheSerPro 700  
Db 2056 CGACACAGCAGCGGGACAGGGTGCAGGCAGATGACCTGGTGAAGATCTCTTCAGCCCA 2115  
Qy 701 GluGluMetAspLeuSerGlnGluGlnProLeuAspAlaGlnGlnGlyProProGluPro 720  
Db 2116 GAAGAGATGGACCTCTCCAGGAGAGAGCCCATGGATGCCAGAGGGGCCCCCGGACCT 2175  
Qy 721 AlaGlnGluSerLeuSerGlySerGluSerLysProLysAspGluLeu 736  
Db 2176 GCACAAGAGTCTCTCTCAGGAGTGAATCGAAGCCCAAGGATGAGCTA 2223  
RESULT 12  
AF097431  
LOCUS Homo sapiens GROSI-S protein mRNA, complete cds.  
DEFINITION Homo sapiens GROSI-S protein mRNA, complete cds.  
ACCESSION AF097431  
VERSION AF097431.1 GI:11127635  
KEYWORDS Homo sapiens (human)  
SOURCE Homo sapiens  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.  
REFERENCE 1 (bases 1 to 2829)  
AUTHORS Kaul, S.C., Sugihara, T., Yoshida, A., Nomura, H. and Wadhwa, R.  
TITLE GroS1, a potential growth suppressor on chromosome 1: its identity

to basement membrane-associated proteoglycan, leprecan  
Oncogene 19 (32), 3576-3583 (2000)  
20406537  
10951563  
2 (bases 1 to 2829)  
Wadhwa, R., Sugihara, T. and Kaul, S.  
Direct Submission  
Submitted (06-OCT-1998) Group 4, Chugai Research Institute of  
Molecular medicine, 153-2 Nagai, Nihari, Ibaraki 300-41, Japan  
Location/Qualifiers  
1..2829  
/organism="Homo sapiens"  
/mol\_type="mRNA"  
/db\_xref="taxon:9606"  
/chromosome="1"  
/tissue\_type="testis"  
52..1143  
/note="member of protein B family"  
/codon\_start=1  
/product="GROSI-S protein"  
/protein\_id="AAG31018.1"  
/db\_xref="GI:11127636"  
/translation="MAVRALKLLTLLAVVAASQAEVESEAGMGVTPDLLFASGTA  
AVARDGPGVLSMERALRSRAALRLRCKCAADFPWELDPWSPSPAQASGAG  
ALRDLSPFGGLLRRAACLRRCIGPPAHSLSSEMELEFRKSPINLYQVAFKINKLE  
KAAVAHTEFFVGNPEHMQNDLYOTMSGKEADFKDLETPHMQPFLGRLYSE  
EOPAVAPFLEALOLEYFVAYEECALCEGPDYDYNLEYNADLFOAITDHYIQVL  
NKKQCVTELASHPSEKPEFDFLPSHYNLQFAYVNIIGNYTQACECAKTYLLFFPND  
EYVQNQLAYAAVLGEHTRSGPREQGT"  
ORIGIN  
Alignment Scores:  
Pred. No.: 3,43e-288 Length: 2829  
Score: 3780.00 Matches: 733  
Percent Similarity: 90.41% Conservative: 2  
Best Local Similarity: 90.16% Mismatches: 1  
Query Match: 97.67% Indels: 77  
DB: Gaps: 2  
US-10-045-815-4 (1-736) x AF097431 (1-2829)  
Qy 1 MetAlaValArgAlaLeuLysLeuThrThrLeuLeuAlaValValAlaAlaSer 20  
Db 52 ATGGCGGTACGCGCGTGAAGCTGTGACACACTGCTGCTGCTGCGCGCTGCTCC 111  
Qy 21 GlnAlaGluValGluSerGluAlaGlyTyrGlyMetValThrProAspLeuPheAla 40  
Db 112 CAAGCCGAGTCCAGTCCGAGGCGAGGATGGGCGCATGTGACGCTCATCTCTCGCC 171  
Qy 41 GluGlyThrAlaAlaTyrAlaArgGlyAspTyrProGlyValValLeuSerMetGluArg 60  
Db 172 GAGGGGACCGACGCTACGCGCGGGGACTGGCCCGGGGTGCTCTGAGCATGGACGG 231  
Qy 61 AlaLeuArgSerArgAlaAlaLeuAlaLeuArgLeuArgCysArgThrGlnCysAla 80  
Db 232 CGCGTGGCTCCCGGGCAGCCCTCCGCGCCCTTCGCTGCGCTGCGCACCCAGTGTGCC 291  
Qy 81 AlaAspPheProTyrGluLeuAspProAspTyrSerProSerProAlaGlnAlaSerGly 100  
Db 292 GCGGACTTCCGCTGGGAGCTGGACCCCGCATGTGTCCCGCCAGCCGCGGCTCGGGC 351  
Qy 101 AlaGlyValAlaLeuArgAspLeuSerPhePheGlyGlyLeuLeuArgArgAlaAlaCysLeu 120  
Db 352 GCGGCGCCCTGCGGACCTGAGCTTCTCGGGGGCTTCTGGGTGCGGTGCTGCTGCTG 411  
Qy 121 ArgArgCysLeuGlyProProAlaAlaHisSerLeuSerGluGluMetGluLeuGluPhe 140  
Db 412 CGCGCTGCTCGGGCGCGCGCGCCCTCGCTCAGCGAAGAGATGGAGCTGGAGTTC 471  
Qy 141 ArgLysArgSerProTyrAsnTyrLeuGlnValAlaTyrPheLysIleAsnLysLeuGlu 160  
Db 472 CGCAAGCGGAGCCCTCAACTACCTCAGGTGCGCTTCTCAAGATCAACAAGTTGGAG 531

JOURNAL  
MEDLINE  
PUBMED  
REFERENCE  
AUTHORS  
TITLE  
JOURNAL  
FEATURES  
source

CDS

QY 161 LysAlaValAlaAlaHisThrPhePheValGlyVasnProGluHisMetGluMetGln 180  
 DB 532 AAAGCTGTTGTCGACGACACACACCTTCTTCGTGGGCAATCCTGAGCACATGGAATGCAG 591  
 QY 181 GlnAsnLeuAspTyrTyrGlnThrMetSerGlyValLysGluAlaAspPheLysAspLeu 200  
 DB 592 CAGAACCTAGACTATTACCAACCAATGCTGTGAGTGAAGAGGCCGACTTCAAGGATCTT 651  
 QY 201 GluThrGlnProHisMetGlnGluPheArgLeuGlyValArgLeuTyrSerGluGluGln 220  
 DB 652 GAGACTCAACCCCATATGCAAGAAATTCGACTGGGAGTGGGCTTACTCAGAGGAACAG 711  
 QY 221 ProGlnGluAlaValProHisLeuGluAlaAlaLeuGlnGluTyrPheValAlaTyrGlu 240  
 DB 712 CCACAGGAAGCTGTGCCCCACCTAGAGGCGGGCTGCAAGAAATCTTTGTGGCTTATGAG 771  
 QY 241 GluCysArgAlaLeuGluGlyProTyrAspTyrAspGlyTyrAsnTyrLeuGluTyr 260  
 DB 772 GAGTGGCGTGGCTCTGCGAAGGCGCTATGCTAGATGGCTACAACTACCTTGGATGAC 831  
 QY 261 AsnAlaAspPheGlnAlaIleThrAspHisTyrIleGlnValLeuAsnCysLysGln 280  
 DB 832 AACGCTGACCTCTCCAGGCCATCACAGATCATTTACATCCAGGTCCTCAACTGTAAAGCAG 891  
 QY 281 AsnCysValThrGluLeuAlaSerHisProSerArgGluLysProPheGluAspPheLeu 300  
 DB 892 AACTGTGTCCAGGAGCTGTCTCCACCAAGTCCAGAGAGGCCCTTTTGAAGACTTCCCTC 951  
 QY 301 ProSerHisTyrAsnTyrLeuGlnPheAlaTyrTyrAsnIleGlyVasnTyrThrGlnAla 320  
 DB 952 CCATCGCATTAATATCTCGAGTTGCTTACTATACATTTGGGAATTTACACAGCT 1011  
 QY 321 GlyGluCysAlaLysThrTyrLeuLeuPhePheProAsnAspGluValMetAsnGlnAsn 340  
 DB 1012 GGTGAATGTCCAGAACCTTATCTCTCTTCTCCCAATACGAGGTGATGACCAAAAT 1071  
 QY 341 LeuAlaTyrTyrAlaAlaMetLeuGlyGluGlnHisThrArgSerIleGlyProArgGlu 360  
 DB 1072 TTGGCTATTATGAGCTATGTTGGAGAGACACACACATCCATCGGCCCGCTGAG 1131  
 QY 361 -----SerAl 362  
 DB 1132 CAGGGCACTAGGAAAGATGTGACCCCGAAAGTACTCAGTTTCCCTCGCCCTGGAGTC 1191  
 QY 362 alysgluTyrArgGlnArgSerLeuLeuGluLysGluLeuLeuPheAlaTyrAspVa 382  
 DB 1192 CAAGAGTACCAGCAGGAGCTTACTGGAAGAAAGACTGCTTTCTTCGCTTATGATGT 1251  
 QY 382 lPheGlyIleProPheValAspProAspSerTrpThrProGluGluValIleProLysAr 402  
 DB 1252 TTTTGGAAATCCCTTTGTGGATCGGATTCATGGACTCCAGAGAAATGATTCCCAAGAA 1311  
 QY 402 gheuGlnGluLysGlnLys----- 408  
 DB 1312 ATTGCAGAGAAACAGAAAGTGAAGACCTTGAAGAAACTGCATGTTGGATCGATCTGATG 1371  
 QY 408 ----- 408  
 DB 1372 AAGCACTTGAGCTTCTTGACCCAGGCAGATGTGAACCTCTGCGAAGGGTGGCGAGGT 1431  
 QY 408 ----- 408  
 DB 1432 CCAGTTTGGGAAGTGGGGTGGAGCCAGGGCTGGCCCTGGAATGCAGTCTCTCAGAGCGG 1491  
 QY 409 -----SerGluArgGluThrAlaValArgIleSerGlnGluIleGlyAsnLe 424  
 DB 1492 TTGTGCTCATAGTCAGAACGGGAAACAGCGCTAGCGATCTCCAGAGATTTGGAACTT 1551  
 QY 424 uMetLysGluIleGluThrLeuValGluGluLysThrLysGluSerLeuAspValSerAr 444  
 DB 1552 TATGAAGGAATCGAGACCCCTTGTGAAGAGAGAACCAAGAGTCACTGGATGTGAGCAG 1611  
 QY 444 gLeuThrArgGluGlyGlyProLeuLeuTyrGluGlyIleSerLeuThrMetAsnSerly 464

DB 1612 ACTGACCGGGAGAGTGGCCCTCTGTATGAAGCATCAGTCTCACCATGACACTCCAA 1671  
 QY 464 sLeuLeuAsnGlyTyrGlnArgValValMetAspGlyValIleSerAspHisGluCysGl 484  
 DB 1672 ACTCTGAAATGGTTACCAGCGGGTGGTGTGAGCGCGCTAAATCTCTGACCAACAGTGTCA 1731  
 QY 484 nGluLeuGlnArgLeuThrAsnValAlaAlaThrSerGlyAspGlyTyrArgGlyGlnTh 504  
 DB 1732 GGAGCTGCAGAGACTGACCAATGTGGCAGCAACTCAGAGATGGCTACCGGGGTGAGAC 1791  
 QY 504 rSerProHisThrProAsnGluLysPheTyrGlyValThrValPheLysAlaLeuLysLe 524  
 DB 1792 CTCCCACTACTCCCAATGAAAAGTTCTATGGTGTCACTGTCTTCAAAGCCCTCAAGCT 1851  
 QY 524 uGlyGlnGluGlyLysValProLeuGlnSerAlaHisLeuTyrTyrAsnValThrGluLy 544  
 DB 1852 GGGGCAAGAGGCAAGATTCTCTGAGAGTGGCCACCTGTACTACAACTGACGAGGAA 1911  
 QY 544 sValArgArgIleMetGluSerTyrPheArgLeuAspThrProLeuTyrPheSerTyrSe 564  
 DB 1912 AGTGGCGGCATCATGAGTCTCTTCCGCTGGATACGCCCTCTACTTTTCTTACTC 1971  
 QY 564 rHisLeuValCysArgThrAlaIleGluValGlnAlaGluArgLysAspAspSerHi 584  
 DB 1972 TCATCTGGTGTGGCGCACTGCCATCGAAGAGGTCCAGGAGAGAGGAGATGATAGTCA 2031  
 QY 584 sProValHisValAspAsnCysIleLeuAsnAlaGluThrLeuValCysValLysGluPr 604  
 DB 2032 TCAGTCCAGCTGGCAACTGATCTCTGATGCCGAGACCTCTGCTGTGTCTCAAGAGCC 2091  
 QY 604 oProAlaTyrThrPheArgAspTyrSerAlaIleLeuTyrLeuAsnGlyAspPheAspGl 624  
 DB 2092 CCCAGCCTACACCTTCGGGCACTACAGGCCATCTCTTACCTAAATGGGCACTTCGATGG 2151  
 QY 624 yGlyAsnPheTyrPheThrGluLeuAspAlaLysThrValThrAlaGluValGlnProGl 644  
 DB 2152 CGAAACTTTTATTTTCTGAACTGGATGCCAAGACCGTGACGGCAGAGGTGACGCTCA 2211  
 QY 644 nCysGlyArgAlaValAlglyPheSerSerGlyThrGluAsnProHisGlyValLysAlaVa 664  
 DB 2212 GTGTGAAGAGCGCTGGGATCTCTTCAGGCCACTGAAAACCCACATGGAGTGAAGCTGT 2271  
 QY 664 lThrArgGlyGlnArgCysAlaAlaIleAlaLeuTyrPheThrLeuAspProArgHisSerGl 684  
 DB 2272 CACAGGGGCGAGCGCTGTGCCATCGCTGTGGTTCACCTGGACCTCGACACAGGGA 2331  
 QY 684 uArgAspArgValGlnAlaAspAlaAspLeuValLysMetLeuPheSerProGluGluMetAs 704  
 DB 2332 CGGGGACAGGGTGCAGGCAGATGACTGTGTGAAGATGCTCTTCAGCCCAAGAGAGATGGA 2391  
 QY 704 pleuSerGlnGluGlnProLeuAspAlaGlnGlnGlyProProGluProAlaGlnGluSe 724  
 DB 2392 CCTCTCCAGGAGAGCCCTCGATGCCAGCAGGGCCCCCCCCCAACCTGCAAGAGTC 2451  
 QY 724 rLeuSerGlySerGluSerLysProLysAspGluLeu 736  
 DB 2452 TCTCTCAGGCATGTGAATCGAAGCCCAAGGATGAGCTA 2488

RESULT 13

AX880254

LOCUS

DEFINITION

AX880254

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

AX880254 Sequence 15159 from Patent EP1074617.

AX880254.1 GI:40034990

Homo sapiens (human)

Homo sapiens

Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

1

Ota, T., Isogai, T., Nishikawa, T., Hayashi, K., Saito, K., Yamamoto, J.,

Ishii, S., Sugiyama, T., Wakamatsu, A., Negai, K. and Otsuki, T.

AX880254 2993 bp DNA linear PAT 17-DEC-2003



TITLE Primers for synthesising full-length cDNA and their use  
JOURNAL Patent: Ep 1074617-A 15159 07-FEB-2001;  
Research Association for Biotechnology (JP)  
FEATURES Location/Qualifiers  
source 1. .2993  
/organism="Homo sapiens"  
/mol\_type="unassigned DNA"  
/db\_xref="taxon:9606"  
CDS 42..2456  
/note="unnamed protein product"  
/codon\_start=1  
/protein\_id="CAE90869.1"  
/db\_xref="GI:40034991"  
/translation="MAVRALKLTLLAVAAASQAESEAGWMTVDLLFAEGTA  
AVARGPWGVLLSMERALSRAALRLRCRTQCAADFPWELDPWSPSPQAAGAA  
ALRDLSPFGLGLRRAALRCLGPPAAHSLSEMELEFRKRSYNYLQVAYFKINLE  
KVAQAATFPVGNPEHMEQNDLYVCTMSGKEADFKLETOPHMQEPRILGVRLYSE  
EQQEAUPELALQYFVAYEBCRALCEGPDYDGYNYLEYNADLFQALTDHYLOVL  
NKQNCVTELAHSPSEKFEYDFPSHNTLQFYNYNIGNYQAVECAKTYLLFPND  
EVNNQNYAALVGEHRSIGTPRESAKEYRQSLLEKELLFFAYDVFIPFVDPDS  
WTPGEVLPKQKQKSERETAVRIISOEINLMKEITELVEEKTESLVSRLTREGG  
PLLYEGISLTMSKLLNGSQRVVMDGVISDHEQBELQRLTNVAATSGDGYRGTSPT  
PNKFGVTVFKALQGOEKVPLQSAHLVYNTVEKVRIMESYFRLDTPLYPSVSHL  
VCTAEVQAEKSDSHPVHVDNCLINAEFLVCVEPPAYTRDYSAILYINGDDPG  
GNFYFELDAKTVAEVQPOCGRAVGFSSGTENPHGVKAVTRQRCALIAFWLTDPRH  
SERVRAARGESNWCDDPPEPFPFAFLFPKSHCQWLRHERSTWDTSNALSLMHC  
LVLPFGPAVNGIQVKEVKGKDSAEFLVPSLGTSAVLFORVGPAGKEMSLGPNRLPC  
PLGSSS"

ORIGIN

Alignment Scores:  
Pred. No.: 6,14e-284 Length: 2993  
Score: 3726.50 Matches: 730  
Percent Similarity: 83.33% Conservative: 0  
Best Local Similarity: 83.33% Mismatches: 6  
Query Match: 96.29% Indels: 141  
Gaps: 1

US-10-045-815-4 (1-736) x AX880254 (1-2993)

Qy 1 MetAlaValArgAlaLeuLysLeuLeuThrThrLeuLeuAlaValValAlaAlaSer 20  
Db 42 ATGCGGTACGCGGTGGAAGCTGCTGACCACACACTGCTGCTGCTGCTGCTGCTC 101  
Qy 21 GluAlaGluValGluSerGluAlaGlyTrpGlyMetValThrProAspLeuPheAla 40  
Db 102 CAAGCCGAGGTCGAGTCGAGCGAGGATGGGGCATGTGTGCTGCTGCTGCTGCTC 161  
Qy 41 GluGlyThrAlaAlaAlaArgGlyAspTrpProGlyValValLeuSerMetGluArg 60  
Db 162 GAGGGGACCGGAGCTACGCGCGCGGGGACTGGCCCGGGGTGGTCTCTGAGCATGGAACG 221  
Qy 61 AlaLeuArgSerArgAlaAlaLeuArgAlaLeuArgLeuArgCysArgThrGlnCysAla 80  
Db 222 GCGCTGCGCTCCCGGGCAGCCCTCCGCGCCCTTCGCTGCTGCTGCTGCTGCTGCTG 281  
Qy 81 AlaAspPheProTrpGluLeuAspProAspTrpSerProSerProAlaGlnAlaSerGly 100  
Db 282 GCCGACTTCCCGTGGAGCTGGACCCGACCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 341  
Qy 101 AlaGlyAlaLeuArgAspLeuSerPhePheGlyGlyLeuLeuArgArgAlaAlaCysLeu 120  
Db 342 GCGCGCGCCCTGCGGACCTGAGCTTCTTTCGGGGGCTTCTGCTGCTGCTGCTGCTGCTG 401  
Qy 121 ArgArgCysLeuGlyProProAlaAlaHisSerLeuSerGluGluMetGluLeuGluPhe 140  
Db 402 GCGCGCTGCTGCGGCGCGCGCCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 461  
Qy 141 ArgLysArgSerProTrpAsnTyrLeuGlnValAlaTyrPheLysIleAsnLysLeuGlu 160  
Db 462 GCGAGCGGAGCCCTTACACTACTGCTGAGTCCGCTACTTCAAGATCAACAAGTTGAG 521  
Qy 161 LysAlaValAlaAlaAlaHisThrPhePheValGlyAsnProGluHisMetGluMetGln 180

Db 522 AAAGCTGTGTGTCAGCAGCACACACTTCTTCTGGGCAATCTGAGCAGCATGGAATGAG 581  
Qy 181 GlnAsnLeuAspTyrTyrGlnThrMetSerGlyValLysGluAlaAspPheLysAspLeu 200  
Db 582 CAGAACCTTAGACTATTACCAACCATGTCTGAGTGAAGAGGCGGACTTCAAGGATCTT 641  
Qy 201 GluThrGlnProHisMetGlnGluPheArgLeuGlyValArgLeuTyrSerGluGluGln 220  
Db 642 GAGACTCAACCCCATATGCAAGAATTTCGACTGGGAGTGCAGTCTACTCAGAGAAACAG 701  
Qy 221 ProGlnGluAlaValProHisLeuGluAlaAlaLeuGlnGluTyrPheValAlaTyrGlu 240  
Db 702 CCACAGGAGCTGTGCCCCACTAGAGCGGGCTGCAAGATACTTTGTGGCTATGAG 761  
Qy 241 GluCysArgAlaLeuCysGluGlyProTyrAspTyrAspGlyTyrAsnTyrLeuGluTyr 260  
Db 762 GAGTCCGCTGCTCTCGGAAGGCGCTATGACTACGATGCTACCACTACCTTCAAGTAC 821  
Qy 261 AsnAlaAspLeuPheGlnAlaIleThrAspHisTyrIleGlnValLeuAsnCysLysGln 280  
Db 822 AACGCTGACCTCTTCCAGGCCATCAGATCATATACATCCAGGTCTCACTGTNAGCAG 881  
Qy 281 AsnCysValThrGluLeuAlaSerHisProSerArgGluLysProPheGluAspPheLeu 300  
Db 882 AACTGTGTGTCAGGAGCTGCTTCCACCCAGTCGAGAGAGCCCTTTTGAAGACTTCTCT 941  
Qy 301 ProSerHisTyrAsnTyrLeuGlnPheAlaTyrTyrAsnIleGlyAsnTyrThrGlnAla 320  
Db 942 CCATGCGATTAATAATTATCTGCACTTCTGCTACTATAACATTTGGGAATTTATACAGGCT 1001  
Qy 321 GlyGluCysAlaLysThrTyrLeuLeuPhePheProAsnAspGluValMetAsnGlnAsn 340  
Db 1002 GTTGAATGTCAGACACTATCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCT 1061  
Qy 341 LeuAlaTyrTyrAlaAlaMetLeuGlyGluGluHisThrArgSerIleGlyProArgGlu 360  
Db 1062 TTGGCTATTATGAGCTATGCTTGGAGAAGAACACACAGATCCATCGCCCGCTGAG 1121  
Qy 361 SerAlaLysGluTyrArgGlnArgSerLeuLeuGluLysGluLeuLeuPhePheAlaTyr 380  
Db 1122 AGTGCCAGAGGATGACGAGGAGGCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCT 1181  
Qy 381 AspValPheGlyIleProPheValAspProAspSerTrpThrProGluGluValIlePro 400  
Db 1182 GATGTTTTTGGAAATCCCTTTTGGATCCGATTCATGGACTCCAGAGAGAGTATCCCTCC 1241  
Qy 401 LysArgLeuGlnGluLysGlnLysSerGluArgGluThrAlaValArgIleSerGlnGlu 420  
Db 1242 AAGAGATTGCAAGAGAAACAGAAAGTCAGAACCGGAAACAGCCGTACGCTATCTCCAGGAG 1301  
Qy 421 IleGlyAsnLeuMetLysGluIleGluThrLeuValGluLysThrLysGluSerLeu 440  
Db 1302 ATTGGGAACCTTATGAGGAATTCGAGACCTTGTGGAAGAGAGAGACCAAGGAGTCACTG 1361  
Qy 441 AspValSerArgLeuThrArgGluGlyProLeuLeuTyrGluGlyIleSerLeuThr 460  
Db 1362 GATGTGAGCAGACTGACCCCGGAAGGTGGCCCTGCTGTATGAGGAGCATCAGTCTCACCC 1421  
Qy 461 MetAsnSerLysLeuLeuAsnGlyTyrGlnArgValValMetAspGlyValIleSerAsp 480  
Db 1422 ATGAATCCAAACTCTCTGAATGTTCCCGCGGTGGTGTGAGCGCGGTAACTCTCTGAC 1481  
Qy 481 HisGluCysGlnGluLeuGlnArgLeuThrAsnValAlaAlaThrSerGlyAspGlyTyr 500  
Db 1482 CACGAGTGTGAGGAGCTGACGAGACTGACCAATGTGGCAGCAACCTCAGGAGATGGCTAC 1541  
Qy 501 ArgGlyGlnThrSerProHisThrProAsnGluLysPheTyrGlyValThrValPheLys 520  
Db 1542 CGGGGTGAGACCTCCCAATGAGGTTCTATGGTGTCTACTGTCTTCAA 1601  
Qy 521 AlaLeuLysLeuGlyGlnGluGlyLysValProLeuGlnSerAlaHisLeuTyrTyrAsn 540



```

Db      1602 GCCTCAAGCTGGGCAAGAGCAAGTTCTCTGAGAGTGCCACCTGTACTACAAAC 1661
QY      541 ValThrGluValArgArgIleMetGluSerTyrPheArgLeuAspThrProLeuTyr 560
Db      1662 GTACGCGAAGAGTGGCGGCATCATGGAGTCTACTTCCGCTGGATACGCCCTCTAC 1721
QY      561 PheSerTyrSerHisLeuValCysArgThrAlaIleGluGluValGluAlaGluArgLys 580
Db      1722 TTTTCTCTACTCTCATCTGTGTGGCGCATCTGCATCGAAGAGGTCCAGCGAGAGGAAG 1781
QY      581 AspAspSerHisProValHisValAspAsnCysIleLeuAsnAlaGluThrLeuValCys 600
Db      1782 GATGATAGTATCATCCAGTCCAGCTGCACAACTGCATCTCTGATCGCGAGACCTCGTGTGT 1841
QY      601 ValLysGluProProAlaTyrThrPheArgAspTyrSerAlaIleLeuTyrLeuAsnGly 620
Db      1842 GTCAAAGAGCCCCCAGCCCTACACCTTCCGCGACTACAGCGCATCTCTTACCTAAATGGG 1901
QY      621 AspPheAspGlyGlyAsnPheTyrPheThrGluLeuAspAlaLysThrValThrAlaGlu 640
Db      1902 GACTTCGATGGCGGAACCTTTATTTCACTGACTGGATGCCAGACCGTACGCGCAGAG 1961
QY      641 ValGlnProGlnCysGlyArgAlaValGlyPheSerSerGlyThrGluAsnProHisGly 660
Db      1962 GTGCAGCCTCAGTGTGAAGAGCCGCTGGGATTTCTTTCAGGCACTGAAAAACCCACATGA 2021
QY      661 ValLysAlaValThrArgGlyGlnArgCysAlaIleAlaLeuTyrPheThrLeuAspPro 680
Db      2022 GTGAAGGCTGTACACAGGGCGAGCGCTGTGCCATCGCCCTGTGTTCCCTCGACCCCT 2081
QY      681 ArgHisSerGluArg----- 685
Db      2082 CGACACAGCGAGCG-GGTGAGAGCAGCTCGAGCGGGTGAGAGCAGCTGGTCTGTGTGTGA 2140
QY      685 ----- 685
Db      2141 CCCGTTCACAGAGCCCTTGGTTGGCTTTCTCTTCCCCAAATCCCATTTGCCAGTGCT 2200
QY      685 ----- 685
Db      2201 GAGACACGAAGAGGACACTTGGACACCAAGCTCCACCGCCCTGTCAATATGGTCACATTG 2260
QY      685 ----- 685
Db      2261 CTTTGTCTCCTCGGGCTGCTGTGAACGGGATCCAGTGCGGGAAGAGAGTCAAGACAGG 2320
QY      685 ----- 685
Db      2321 GAGCGATGCTGAGTCTTGTGTTCTCTCTGGGCCCCCACTTCAGCTGTCTCTTTCCAGAG 2380
QY      685 ----- 685
Db      2381 AGTAGGACCTGCTGGAAGGAGATGAGCCTGGGGCCATTAAAGAAACCTTCTTGTCTCCCT 2440
QY      685 ----- 685
Db      2441 GGGAGTAGCAGCTGAGAGATAGCGAGTGTCTGGAGCGGAGGCTCTCTGAATGGGGCAGG 2500
QY      686 ----- AspArgValGlnAlaAspAspLeuValLysMetLeuPheSerPro 700
Db      2501 GGTTTGTCTTGTGAGGACAGGGTGCAGGAGATGACCTGTGTGAAGATGCTCTTTCAAGCCA 2560
QY      701 GluGluMetAspLeuSerGlnGluProLeuAspAlaGlnGlyProProGluPro 720
Db      2561 GAAGAGATGTTCTCTCCAGAGACGCCCTGTGATGCCAGAGGGCCCCCGAACCT 2620
QY      721 AlaGlnGluSerLeuSerGlySerGluSerLysProLysAspGluLeu 736
Db      2621 GCACAAGATCTCTCTCAGGACAGTGAATCGAAGCCCAAGGATGAGCTA 2668

```

RESULT 14  
BD158280  
LOCUS

DEFINITION  
ACCESSION  
VERSION  
KEYWORDS  
SOURCE  
ORGANISM

Primer for synthesizing full-length cDNA and use thereof.

BD158280

BD158280.1 GI:27864038

JP 2002191363-A/13123

Homo sapiens (human)

Homo sapiens

Bukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

(bases 1 to 2993)

Oca, T., Isogai, T., Nishikawa, T., Hayashi, K., Saito, K., Yamamoto, J.,

Ishii, S., Sugiyama, T., Wakamatsu, A., Nagai, K. and Otsuki, T.

Primer for synthesizing full-length cDNA and use thereof

Patent: JP 2002191363-A 13123 09-JUL-2002;

HELEX RESEARCH INSTITUTE

CS Homo sapiens (human)

PN JP 2002191363-A/13123

PD 09-JUL-2002

PF 28-JUL-2000 JP 2000280990

PI TOSHIO OTA, TAKAO ISOGAI, TETSUO NISHIKAWA, KOJI HAYASHI, KAORU

PI SAITO,

PI JUNICHI YAMAMOTO, SHIZUKO ISHII, TOMOYASU SUGIYAMA, AI WAKAMATSU,

PI KEIICHI NAGAI, TETSUJI OTSUKI

PC

C12N15/09, C07K14/47, C07K16/18, C12N1/15, C12N1/19, C12N1/21, C12N5/ PC

10,

PC, C12P21/02, C12Q1/68//C12P21/08, G06F17/30, C12N15/00, C12N5/00 CC

Primer for synthesizing full-length cDNA and use thereof FH Key

Location/Qualifiers

(42)..(2453).

FEATURES

source

1..2993

/organism="Homo sapiens"

/mol\_type="genomic DNA"

/db\_xref="taxon:9606"

ORIGIN

Alignment Scores:

Pred. No.: 6,14e-284 Length: 2993

Score: 3726.50 Matches: 730

Percent Similarity: 83.33% Conservative: 0

Best Local Similarity: 83.33% Mismatches: 6

Query Match: 96.29% Indels: 141

DB: Gaps: 1

US-10-045-815-4 (1-736) x BD158280 (1-2993)

QY 1 MetAlaValArgAlaLeuLysLeuLeuThrThrLeuAlaValAlaAlaAlaSer 20

Db 42 ATGGCGGTACGCGCGCTGAAGCTGTGACACACTGTGGCTGTGTCGGCGCTGCCCTCC 101

QY 21 GlnAlaGluValGluSerGluAlaGlyTyrGlyMetValThrProAspLeuPheAla 40

Db 102 CAAGCCGAGTCTCCGAGGACGATGGGCGATGGTGGCTGTGTCCTTCTTCC 161

QY 41 GluGlyThrAlaAlaTyrAlaArgGlyAspTyrProGlyValValLeuSerMetGluArg 60

Db 162 GAGGGGACCGGACGCTACGCGCGGGGACTGGCCCGGGGTGGTCTTGAGCATGGAAACGG 221

QY 61 AlaLeuArgSerArgAlaAlaLeuArgAlaLeuArgLeuArgCysArgThrGlnCysAla 80

Db 222 GCGCTGCGCTCCCGGCGAGCCCTCCGCGCCCTTCCGCTGCGCTGCGCGACCATGTGTGCC 281

QY 81 AlaAspPheProTyrGluLeuAspProAspTyrPsrProSerProAlaGlnAlaSerGly 100

Db 282 GCGGACTTCCGCTGGGAGCTGGACCCGACTGTGTCCCGCGCCCGCCAGCCCTCGGGC 341

QY 101 AlaGlyAlaLeuArgAspLeuSerPheGlyGlyLeuLeuArgArgAlaAlaCysLeu 120

Db 342 GCGCGCGCCCTTCCGCGACCTGAGGTCTTCCGCGGCGCTTCTGCGCTGCGCTGCGCTG 401

QY 121 ArgArgCysLeuGlyProProAlaAlaHisSerLeuSerGluMetGluLeuGluPhe 140

Db 402 GCGCGCTCTCGGCGCGCCCGCCACTCGCTCAGCGAAGAGATGGAGCTGGAGTTC 461

141 ArgLysArgSerProTyrAsnTyrLeuGlnValAlaTyrPheLysIleAsnLysLeuGlu 160  
142 CGCAGCGGAGCCCTACAACTACCTCGAGTCCGCTACTTCAAGATCAACAAGTTGGAG 521  
161 LysAlaValAlaAlaAlaHisThrPhePheValGlyAsnProGluHisMetGluMetGln 180  
162 AAAGCTGTGTGTCAGACACACACCTTCTTCTGCGCAATCTCTGAGCAGCATGGAAATGCAG 581  
181 GlnAsnLeuAspTyrTyrGlnThrMetSerGlyValLysGluAlaAspPheLysAspLeu 200  
182 CAGAACTAGACTATTACCAACCATGTCTGGAGTGAAGAGGCCGACTTCAAGATCTT 641  
201 GluThrGlnProHisMetGlnGluPheArgLeuGlyValArgLeuTyrSerGluGluGln 220  
202 GAGACTCAACCCCATATGCAAGAAATTCGACTGGAGTGGCACTACTCAGAGGAACAG 701  
221 ProGlnGluAlaValProHisLeuGluAlaLeuGlnGluTyrPheValAlaTyrGlu 240  
202 CCACAGAGAGCTGTGCCCCACCTAGAGCGCGCTGCAAGAAATCTTTGTGGCTATGAG 761  
241 GluCysArgAlaLeuCysGluGlyProTyrAspTyrAspGlyTyrAsnTyrLeuGluTyr 260  
242 GAGTGGCGTGCCTCTCGGAAGGCGCTATGACTACGATGGCTACAACTACCTTGAGTAC 821  
261 AsnAlaAspLeuPheGlnAlaIleThrAspHisTyrIleGlnValLeuAsnCysLysGln 280  
262 AACGCTGACCTCTTCCAGGCCATCACAGATCATTTACATCCAGGTCTTCAACTGTGAACAG 881  
281 AsnCysValThrGluLeuAlaSerHisProSerArgGluLysProPheGluAspPheLeu 300  
282 AACTGTGTACGAGCTGTCTTCCACCCCAAGTGCAGAGAGCCCTTTGAAGACTTCTC 941  
301 ProSerHisTyrAsnTyrLeuGlnPheAlaTyrTyrAsnIleGlyAsnTyrThrGlnAla 320  
302 CCATCGCATTAATATCTCGATTTGCCTACTATAACATTTGGGAATTTATACACAGCT 1001  
321 GlyGluCysAlaLysThrTyrLeuLeuPhePheProAsnAspGluValMetAsnGlnAsn 340  
1002 GTTGAATGTGCAGAGACCTACTTCTCTTCTTCCCAATGACAGGTGATGAACCAAAAT 1061  
341 LeuAlaTyrTyrAlaAlaMetLeuGlyGluHisThrArgSerIleGlyProArgGlu 360  
1062 TTGGCTATTATGTCAGCTATCTTGGAGAGAAACACACACAGATCCATCGGCCCGTGAG 1121  
361 SerAlaLysGluTyrArgGlnArgSerLeuLeuLysGluLeuLeuPheAlaTyr 380  
1122 AGTGCCAAAGGAGTACCGACAGCGAGCTACTGGAAAAAGAACTGCTTTCTTCGCTTAT 1181  
381 AspValPheGlyIleProPheValAspProAspSerTyrThrProGluGluValIlePro 400  
1182 GATGTTTTTGAATTCCTTTGTGGATCCGGATTCATGGACTCCAGGAGAGTCAATCCC 1241  
401 LysArgLeuGlnGluLysGlnLysSerGluArgGluThrAlaValArgIleSerGlnGlu 420  
1242 AAGAGATTGCAAGAGAAACAGAAAGTCAGACGGGAAACAGCGCTACGCTATCCAGGAG 1301  
421 IleGlyAsnLeuMetLysGluIleGluThrLeuValGluGluLysThrLysGluSerLeu 440  
1302 ATTGGGAACCTTATGAAGAAATCAGACCTTTGTGGAAAGAGAAACCAAGAGTCACTG 1361  
441 AspValSerArgLeuThrArgGluGlyGlyProLeuLeuTyrGluGlyIleSerLeuThr 460  
1362 GATGTGACGACACTGACCCGGAGAGGTGGCCCTCTGTGTATGAAGGCATCAGTCTCACC 1421  
461 MetAsnSerLysLeuLeuAsnGlyTyrGlnArgValValMetAspGlyValIleSerAsp 480  
1422 ATGAATCCAACTCTGAATGGTTCCCGAGCGGGGTGGTGGAGCGCGTAAATCTCTGAC 1481  
481 HisGluCysGlnGluLeuGlnArgLeuThrAsnValAlaAlaThrSerGlyAspGlyTyr 500  
1482 CACGAGTGTACGAGACTGCAGAGACTGACCAATGTGGCAGCAACCTCAGGAGATGGCTAC 1541

501 ArgGlyGlnThrSerProHisThrProAsnGluLysPheTyrGlyValThrValPheLys 520  
1542 CGGGGTTCAGACCTCCCACTATCCCAATGAAAGTTCTATGGTGTCACTGTCTTCAAA 1601  
521 AlaLeuLysLeuGlyGlnGluLysValProLeuGlnSerAlaHisLeuTyrTyrAsn 540  
1602 GCCCTCAAGCTGGGCAAGAGGCAAGTTCTCTGCAGAGTGCCCACTGTACTACAAAC 1661  
541 ValThrGluLysValArgArgIleMetGluSerTyrPheArgLeuAspThrProLeuTyr 560  
1662 GTGACGAGAGGTGGCGCATCATGGAGTCTACTTCCGCTGGATACGCCCTCTACTAC 1721  
561 PheSerTyrSerHisLeuValCysArgThrAlaIleGluGluValGlnAlaGluArgLys 580  
1722 TTTTCTCTACTCTCATCTGTGTGGCGCACTGCCATCGAAGAGTCCAGGAGAGAGAG 1781  
581 AspAspSerHisProValHisValAspAsnCysIleLeuAsnAlaGluThrLeuValCys 600  
1782 GATGATGATCATCTAGCTCCACGCTGGACAACATCATCTCTGAATGCCGAGACCTCGTGT 1841  
601 ValLysGluProProAlaTyrThrPheArgAspTyrSerAlaIleLeuTyrLeuAsnGly 620  
1842 GTCAAGAGCCCGCCAGCTTACACCTTCCGCGACTACAGCGCCATCTTTTACCTAAATGG 1901  
621 AspPheAspGlyGlyAsnPheTyrPheThrGluLeuAspAlaLysThrValThrAlaGlu 640  
1902 GACTTCGATGGCGAAACCTTTTACTGAATGGATGCCAAGACCGTGCAGCGAGAG 1961  
641 ValGlnProGlnCysGlyArgAlaValGlyPheSerSerGlyThrGluAsnProHisGly 660  
1962 GTGCACCTCAGTGTGAAGAGCCGTGGATTTCTCTTACGGCACTGAAACCCACATGGA 2021  
661 ValLysAlaValThrArgGlyGlnArgCysAlaIleAlaLeuTyrPheThrLeuAspPro 680  
2022 GTGAAGCTGTCAACAGGGGCGAGCGCTGTGCATCGCCCTGTGGTTCACTTCCCTGACCT 2081  
681 ArgHisSerGluArg----- 685  
2082 CGACACAGCGAGCG-GGTGAGAGAGCTCGAGCGGTGAGAGCGTGGTGTGTGTGA 2140  
685 ----- 685  
2141 CCGTTCCCGAGAGCGCCCTTGTGTTGCCCTTCTTCTTCCCAATCCATTGCGAGTGGCT 2200  
685 ----- 685  
2201 GAGACAGAAAGAGGACACTTGGGACACAGCTCCACGCGCTGTCAATTATGTCACATTG 2260  
685 ----- 685  
2261 CCTGTCTCCTCCTGGCCCTGCTGTGAACGGGATCCAGSTGGGAAAGAGGTCAAGACAGG 2320  
685 ----- 685  
2321 GAGCGATGTGAGTTCTTGTTCCTTCCCTTGGCCCCCACTTCAGCTGTCTTTTCCAGAG 2380  
685 ----- 685  
2381 AGTAGGACCTGTGGCAAGAGGATGAGCTGGGGCCATTAAAGAACTTCTTGTCCCT 2440  
685 ----- 685  
2441 GGGAGTAGCAGCTGAGAGATACCGAGTGTCTGGAGCGGAGGCTCTCTGAATGGGAGG 2500  
686 ----- 686  
2501 GGTTTGTCTTGCAGGACAGGGTGCAGGAGATGACCTGTGTGAAGATGCTCTTCAGCCCA 2560  
701 GluGluMetAspLeuSerGlnGluGlnProLeuAspAlaGlnGlnGlyProProGluPro 720  
2561 GAAGAGATGTTCTCTCCAGGAGCAGCCCTGGATGCCAGCGGGCCCCCGGAACCT 2620  
721 AlaGlnGluSerLeuSerGlySerGluSerLysProLysAspGluLeu 736

```

Db      2621 GCACAGAGTCTCTCTCAGGAGTGAATCGAAGCCCAAGGATGAGCTA 2668
|||||
RESULT 15
LOCUS   AK027648
DEFINITION Homo sapiens cDNA FLJ14742 fis, clone NT2RP3002650, highly similar
to Mus musculus growth suppressor 1L (Gros1) mRNA.
ACCESSION AK027648
VERSION   GI:14042480
KEYWORDS  oligo capping; fis (full insert sequence).
SOURCE    Homo sapiens
ORGANISM  Homo sapiens
REFERENCE 1
AUTHORS   Isogai,T., Ota,T., Hayashi,K., Sugiyama,T., Otsuki,T., Suzuki,Y.,
Nishikawa,T., Nagai,K., Sugano,S., Shiratori,A., Sudo,H.,
Wagatsuma,M., Hosoi,T., Kaku,Y., Kodaira,H., Kondo,H.,
Sugawara,M., Takahashi,M., Chiba,Y., Ishida,S., Murakawa,K.,
Ono,Y., Takiguchi,S., Watanabe,S., Kimura,K., Murakami,K.,
Ishii,S., Kawai,Y., Saito,K., Yamamoto,J., Wakamatsu,A.,
Nakamura,Y., Nagahara,K., Masuho,Y., Ninomiya,K. and Iwayanagi,T.
NEDO human cDNA sequencing project
Unpublished
2 (bases 1 to 2993)
Isogai,T. and Otsuki,T.
Direct Submission
Submitted (10-MAY-2001) Takao Isogai, Helix Research Institute,
Genomics Laboratory; 1532-3 Yana, Kisarazu, Chiba 292-0812, Japan
(E-mail:genomics@helix.co.jp, Tel:81-438-52-3975, Fax:81-438-52-3986)
NEDO human cDNA sequencing project supported by Ministry of
Economy, Trade and Industry of Japan; cDNA full insert sequencing:
Research Association for Biotechnology; cDNA library construction;
5'- & 3'-end one pass sequencing and clone selection; Helix
Research Institute (supported by Japan Key Technology Center etc.)
and Department of Virology, Institute of Medical Science,
University of Tokyo.
FEATURES
Location/Qualifiers
1..2993
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="NT2RP3002650"
/cell_lines="NT2"
/cell_type="teratocarcinoma"
/clone_lib="NT2RP3"
/notes="Cloning vector: pME18SPL3-mRNA from NT2 neuronal
precursor cells after 2-weeks retinoic acid (RA)
induction."
42..2456
/notes="unnamed protein product"
/codon_start=1
/protein_id="BAB55264.1"
/db_xref="GI:14042481"
/translation="MAVRALKLTLTLLAVVAASQAQVESEAGWMVTPDLLFAEGTA
AYARDGFWVLNMRALRRALRLRCRTQCAADFPEWLDPPDWSPPQAASGAA
ALRDLSPFGGLRPAACLRCLGPPAAHSISEMELEFRKRSFYNTLQVAFYKINLE
KAVAAHTFFVGNPEHEMOMQNLDDYOTMSGVKEADPKOLETQPHMOEFLRGLYLYSE
EPOEAVPHLEALQOEYFVAYEBCRALCEGPDYDGVNLYENADLFOAITDHYIOLV
NCKNVCYTELASHESREKPEDFLPSHYNLYQFAYNIGNVTOAVCAKTYLLFFND
EVNQNLYAVAAMLGEHRTSIRIGPRESAKERYORSLLLEKLLFFADVGVIPVDGDS
WTPGEVLPKRLQEKQERSTAVRISOEIGNLWKEIEITLVEETKESLDVSRUTREGG
PLLYEGSLVTNWSKLNGSORVNDGVISDHECOELQRLNVAATSGDVRGQTSPT
NKEKFGYVTFKALKLQEGKVPQLQSAHLYNTEKVRIMESYFFRLDPLFYFSYGH
VCFRTAIEVQAEKRDSDHPVNDNCILNAETLVCVKEPPAYTRFDYSAILYNGDDPG
GNFYFTBEDAKVTAEVQPOCGRAVGPSSGTENPHGVKAVTRGORCAIALWFLDPRH
SERVRAARAGESWCCGDPFPERFWFAFLPKSHCQWLHERSTWDTSSNLSLWSHC
LVLPGPVANGIQGVKEVKTGSDAEFLVPSLIGPTSAVLFRQVGPAGKEMSLGPLRLNLP
PLGSSS"
ORIGIN
Alignment Scores:

```

```

Pred. No.: 6,14e-284 Length: 2993
Score: 3726.50 Matches: 730
Percent Similarity: 83.33% Conservative: 6
Best Local Similarity: 83.33% Mismatches: 6
Query Match: 96.29% Indels: 141
DB: 5 Gaps: 1
US-10-045-815-4 (1-736) x AK027648 (1-2993)
Qy 1 MetAlaValArgAlaLeuLysLeuLeuThrThrLeuLeuAlaValValAlaAlaLaser 20
Db 42 ATGGCGGTACGCGGTGAAGTGTGACACACACTGCTGGCTGTCTGGCCGCGCTGCCCTCC 101
Qy 21 GlnAlaGluValGluSerGluAlaGlyTTPGlyMetValThrProAspLeuLeuPheAla 40
Db 102 CAAGCCGAGTCCAGTCCGAGGAGGATGGGCGATGGTGCAGCGCTGATCTGCTCTCGGC 161
Qy 41 GluGlyThrAlaAlaValAlaArgGlyAspTTPProGlyValValLeuSerValGluArg 60
Db 162 GAGGGACCGACGCTTACCGCGCGCGGGAGTGGCCCGGGTGGTCTGTGAGCATGGAACGG 221
Qy 61 AlaLeuArgSerArgAlaAlaLeuArgAlaLeuArgGlyLeuArgCysArgThrGlnCysAla 80
Db 222 GCGCTCGCTCCCGGCGACGCTCCGCGCCCTTCGCTCGGCTGGCGCACCCAGTGTGCC 281
Qy 81 AlaAspPheProTTPGluLeuAspProAspTTPSerProSerProAlaGlnAlaSerGly 100
Db 282 GCGGACTTCCGTTGGGAGTGGACCCCGACTCGTCCCCAGCCCGCCGAGCCCTCGGGC 341
Qy 101 AlaGlyAlaLeuArgAspLeuSerPhePheGlyGlyLeuLeuArgArgAlaAlaCysLeu 120
Db 342 GCGCGCCCTCGCGGACCTGAGCTTCTTCGGGGGCGCTTCTGCGTGGCGCTGCGCTG 401
Qy 121 ArgArgCysLeuGlyProProAlaAlaHisSerLeuSerGluGluMetGluLeuPhe 140
Db 402 CGCGCTGCTCGCGCGCGCGCGCCGCGCCACTCGCTCAGCGAAGAGATGGAGTGTTC 461
Qy 141 ArgLysArgSerProTTPAsnTyrLeuGlnValAlaTyrPheLysIleAsnLysLeuGlu 160
Db 462 CGCAGCGGAGCGCCCTACAACTACTTGGAGGTGGCTACTTCAAGATCAACAGTTGGAG 521
Qy 161 LysAlaValAlaAlaAlaHisThrPhePheValGlyAsnProGluHisMetGluMetGln 180
Db 522 AAAGCTGTGTCTGCAGACACACACACTTCTTGGTGGGCAATCTCTGAGCACATGGAATGCAG 581
Qy 181 GlnAsnLeuAspTyrTyrGlnThrMetSerGlyValLysGluAlaAspPheLysAspLeu 200
Db 582 CAGACCTTAGACTATTACCAACCATGTCTGGAGTGAAGGAGCGCGACTTCAAGGATCTT 641
Qy 201 GluThrGlnProHisMetGlnGluPheArgLeuGlyValArgLeuTyrSerGluGln 220
Db 642 GAGACTCAACCCCATATGCAAGATTTTCGACTGGGAGTGGGACTCTACTCAGAGGAACAG 701
Qy 221 ProGlnGluAlaValProHisLeuGluAlaLeuGlnGluTyrPheValAlaTyrGlu 240
Db 702 CCACAGGAAGCTGTGCCCCCAGCTAGAGCGCGCTGCAAGAAATATCTTGTGGCTATGAG 761
Qy 241 GluCysArgAlaLeuCysGluGlyProTyrAspTyrAspGlyTyrAsnTyrLeuGluTyr 260
Db 762 GAGTGGCGTGCCTCTGCGAAGGGGCCCTATGACTACGATGGCTACACTACTTGGAGTAC 821
Qy 261 AsnAlaAspLeuPheGluAlaIleThrAspHisTyrIleGlnValLeuAsnCysLysGln 280
Db 822 AACGCTGACCTTCCAGCGCCATCATCAGATCATTTACATCCAGGCTCCTCACTGTAAGCAG 881
Qy 281 AsnCysValThrGluLeuAlaSerHisProSerArgGluLysProPheGluAspPheLeu 300
Db 882 AACTGTGTCGAGGCTTGTCTCCACCCCAAGTCGAGAGAGAGCCCTTTGAAGACTTCTC 941
Qy 301 ProSerHisTyrAsnTyrLeuGlnPheAlaTyrTyrAsnIleGlyAsnTyrThrGlnAla 320
Db 942 CCATCGCATTATAATTATCTGAGTTTSCCTACTATAACATTGGGAATTTATACAGGCT 1001

```

QY 321 GlyGluCysAlaLysThrTyrLeuLeuPhePheProAsnAspGluValMetAsnGlnAsn 340  
Db 1002 GTTGAATGTCCAAAGACCTATCTCTCTCTCCCAATGACGAGGTGATGAACCAAAAT 1061  
QY 341 LeuAlaTyrTyrAlaAlaMetLeuGlyGluGluHisThrArgSerIleGlyProArgGlu 360  
Db 1062 TTGGCCTATTATGAGTATGCTTGTGAGAAAGACACACAGATCCATCGCCCGGTGAG 1121  
QY 361 SerAlaLysGluTyrArgGlnArgSerLeuLeuGluLysGluLeuLeuPheAlaTyr 380  
Db 1122 AGTGCCAAAGAGTACCGACAGCGAAGCCTACTGGAAGAAAGAACTGCTTTCTTCGCTTAT 1181  
QY 381 AspValPheGlyIleProPheValAspProAspSerTyrProGluGluValIlePro 400  
Db 1182 GATGTTTTTGGAAATCCCTTTGTGGATCCGGATTCATGGACTCCAGGAGAAGTATCC 1241  
QY 401 LysArgLeuGlnGluLysGlnLysSerGluArgGluThrAlaValArgIleSerGlnGlu 420  
Db 1242 AAGAGATTGCAGAGAAACAGAGTCAAGACCGGAAACACACCGTACGCACTCTCCAGGAG 1301  
QY 421 IleGlyAsnLeuMetLysGluIleGluThrLeuValGluLysThrLysGluSerLeu 440  
Db 1302 ATTGGGAACCTTATGAAGGAATCCAGACCCCTTGTGGAAGAGAAGACCAAGGAGTCACTG 1361  
QY 441 AspValSerArgLeuThrArgGluGlyGlyProLeuLeuTyrGluGlyIleSerLeuThr 460  
Db 1362 GATGTGAGCACTGACCCGGGAAGGTGGCCCCCTGCTGTATGAGGCATCAGTCTCACC 1421  
QY 461 MetAsnSerLysLeuLeuAsnGlyTyrGlnArgValValMetAspGlyValIleSerAsp 480  
Db 1422 ATGAACTCCAACTCCTGAATGGTTCCAGCGGGTGGTGTGAGCGCGCTAATCTCTGAC 1481  
QY 481 HisGluCysGlnGluLeuGlnArgLeuThrAsnValAlaAlaThrSerGlyAspGlyTyr 500  
Db 1482 CACGAGTGTACGAGAGTGCAGAGACTGACCAATGTGCAGCAACCTCAGGAGATGGCTAC 1541  
QY 501 ArgGlyGlnThrSerProHisThrProAsnGluLysPheTyrGlyValThrValPheLys 520  
Db 1542 CGGGGTACAGACTCCCCACACTCTCCCAATGAAAGTTCTATGTGTCTCTTCAAA 1601  
QY 521 AlaLeuLysLeuGlyGlnGluLysValProLeuGlnSerAlaHisLeuTyrTyrAsn 540  
Db 1602 GCCCTCAAGCTGGGCAAGAAAGCAAAAGTTCTCTGACAGAGTGGCCACCTGTACTACAAC 1661  
QY 541 ValThrGluLysValArgIleMetGluSerTyrPheArgLeuAspThrProLeuTyr 560  
Db 1662 GTGACGGAGAAGGTGCGCGCATCATGAGAGTCTACTTCCGCTGGATACGCCCTCTAC 1721  
QY 561 PheSerTyrSerHisLeuValCysArgThrAlaIleGluGluValGlnAlaGluArgLys 580  
Db 1722 TTTTCTTACTCTCATCTGGTGTGCGCACTGCCATCGAAGAGGTCCAGGCAGAGGAAG 1781  
QY 581 AspAspSerHisProValHisValAspAsnCysIleLeuAsnAlaGluThrLeuValCys 600  
Db 1782 GATGATAGTCAATCCAGTCCACGTGGACAACCTGCATCTCGAAGAGGTCCAGGCAGAGGAAG 1841  
QY 601 ValLysGluProProAlaTyrThrPheArgAspTyrSerAlaIleLeuTyrLeuAsnGly 620  
Db 1842 GTCAAGAGCCCCCAGCCCTACACCTTCGCGACTACAGGCCATCTTTACCTAAATGGG 1901  
QY 621 AspPheAspGlyGlyAsnPheTyrPheThrGluLeuAspAlaLysThrValThrAlaGlu 640  
Db 1902 GACTTCGATGGCGGAAACTTTTATTTACTGAACCTGATGCCAAGACCGTGACGGCAGAG 1961  
QY 641 ValGlnProGlnCysGlyArgAlaValGlyPheSerSerGlyThrGluAsnProHisGly 660  
Db 1962 GTGCAGCCCTCAGTGTGGAAGAGCCGTGGGATTTCTTTACGGACCTGAAACCCCATGGA 2021  
QY 661 ValLysAlaValThrArgGlyGlnArgCysAlaIleAlaLeuTyrPheThrLeuAspPro 680  
Db 2022 GTGAGGCTGTCAACAGGGGGCAGGCGCTGTGCCCTGTGCTTACCCCTGACCCCT 2081  
QY 681 ArgHisSerGluArg----- 685

Db 2082 CGACACAGGAGCG--GGTGAGAGCAGCTCGAGCGGTGAGAGCAGCTGGTCTGTGTGA 2140  
QY 685 ----- 685  
Db 2141 CCGGTTCCAGAGCGCCCTTGTGTTTTCCTTCTTCCCAAATCCCAATCCAGTGCGT 2200  
QY 685 ----- 685  
Db 2201 GAGACACGAAAGAGGACACTTGGGACACCCAGCTCCACGCCCTGTCTATTATGGTCACATTG 2260  
QY 685 ----- 685  
Db 2261 CCTTGTCTCCCTGGGCGCTGTGTGAACGGGATCCAGGTGGGAAAGAGGTCAAGACAGG 2320  
QY 685 ----- 685  
Db 2321 GAGCGATGCTGAGTTCTTGTTCCTCTTGGGCCCCACTTCAGCTGTCTTTTCCAGAG 2380  
QY 685 ----- 685  
Db 2381 AGTAGACCTGTCTGGGAAGAGATGAGCCTGGGGCCATTAAAGGAACCTTTCCTTGTCCCT 2440  
QY 685 ----- 685  
Db 2441 GGGAAAGTAGCAGCTGAGAGATAGCGAGTGTCTGGAGCGGAGCGCTCTCTGAATGGCAGG 2500  
QY 686 -----AspArgValGlnAlaAspAspLeuValLysMetLeuPheSerPro 700  
Db 2501 GGTTCCTCCTTCAGGACAGGGTGCAGGAGATGACCTGGTGAAGATGCTCTTCAGCCCA 2560  
QY 701 GluGluMetAspLeuSerGlnGluGlnProLeuAspAlaGlnGlnGlyProProGluPro 720  
Db 2561 GAAGAGATGCTCTCTCTCCAGGAGCGCCCTGGATGCCAGCAGGGCCCCCGGAACCT 2620  
QY 721 AlaGlnGluSerLeuSerGlySerGluSerLysProLysAspGluLeu 736  
Db 2621 GCACAAGAGTCTCTCTCAGGCAGTGAATCGAAGCCCAAGATGAGCTA 2668

Search completed: July 18, 2004, 16:08:33  
Job time : 20985 secs

This Page Blank (uspto)

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

Protein - nucleic search, using frame\_plus\_p2n model

Run on: July 18, 2004, 10:01:05 ; Search time 375 Seconds  
(without alignments)  
1089.184 Million cell updates/sec

Title: US-10-045-815-4  
Perfect score: 3870  
Sequence: 1 MAVRALKLLTLLAVAAAS.....PPEPAQESLSGSESKPKDEL 736

Scoring table:

BLOSUM62	
Xgapop 10.0 , Xgapext 0.5	
Ygapop 10.0 , Ygapext 0.5	
Fgapop 6.0 , Fgapext 7.0	
Delop 6.0 , Delext 7.0	

Searched: 682709 seqs, 277475446 residues

Total number of hits satisfying chosen parameters: 1365418

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

```
USER=US10045815 @CN_1_1_128 @runat_14072004_123009_16842 -NCPU=6 -ICPU=3  
-MODE=LOCAL -OUTFWT=ptio -NORX=ext -HEARSIZE=500 -MINLEN=0 -MAXLEN=2000000000  
-LIST=45 -DCCALIGN=200 -THR SCORE=pct -THR MAX=100 -THR MIN=0 -ALIGNA=15  
LOOPTXT=7 -GAPCOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7  
FVGAPEXT=7 -YGAPOPOP=10 -YGAPEXT=10 -XGAPOPOP=10 -XGAPEXT=0.5 -FGAPOPOP=6  
DEV TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPOPOP=10 -XGAPEXT=0.5 -FGAPOPOP=6
```

Result No.	Score	Query		ID	Description	
		Match	Length			
C 1	194.5	5.0	301	4	US-09-439-313-271	Sequence 271, App
C 2	194.5	5.0	301	4	US-09-353-616A-271	Sequence 271, App
C 3	194.5	5.0	301	4	US-09-233-149A-271	Sequence 271, App
C 4	194.5	5.0	301	4	US-09-153-812-271	Sequence 271, App
C 5	194.5	5.0	301	4	US-09-636-215-271	Sequence 271, App
C 6	194.5	5.0	301	4	US-09-685-166A-271	Sequence 271, App
C 7	194.5	5.0	301	4	US-09-688-489-271	Sequence 271, App
C 8	162.5	4.2	1230025	4	US-09-198-452A-1	Sequence 1, Appli
C 9	122.5	3.2	2595	4	US-09-738-946-5	Sequence 5, Appli
C 10	118	3.0	1512	4	US-09-134-001C-643	Sequence 643, App
C 11	116	3.0	1662	4	US-09-134-000C-666	Sequence 666, App
C 12	115.5	3.0	32798	4	US-09-604-634B-1	Sequence 1, Appli

```
Percent Similarity: 61.39% Conservative: 14
Best Local Similarity: 47.52% Mismatches: 17
Query Match: 5.03% Indels: 23
DB: 4 Gaps: 3

US-10-045-815-4 (1-736) x US-09-439-313-271 (1-301)

QY 637 ValThrAlaGluValGln-ProGlnCysGlyArgAlaValGlyPheSerSerGlyThrGl 656
Db 301 GTGACTGCTTCTATAAAACCCAAATGTGGCGCATGATGATTN-TCATCTGGAGGAGA 243
QY 656 uAsnProHisGlyValLysAlaValThrArgGlyGlnArgCysAlaIleAlaLeuTpph 676
Db 242 GAACCNVTCATGGGTGTAAGGAGTCCACCAAGGAAGAGGTTGCTGTGGCTCTGTGGTT 183
QY 676 eThrLeuAspProArgHisSerGluArgAspArgValGlnAlaAspLeuValLysMe 696
Db 182 CACCTTGGACCCACTTTATAGAGAAATGGAGCAATACAGGCTGATGAAGTGAATGCAAT 123
QY 696 tLeuPheSerProGluGluMetAspLeuSerGlnGluGlnProLeuAspAlaGlnGlnGl 716
Db 122 T-----CTGGATCAAGAA-----CAGCAAGG 102
QY 716 yProProGluProAlaGlnGluSerLeuSerGlySerGluSerLysProLysAspGluLe 736
Db 101 GAAGCATGAA-----CTGAATATCAACCCCTAAAGATGAGCT 66
QY 736 u 736
Db 65 A 65

RESULT 2
US-09-352-616A-271/c
; Sequence 271, Application US/09352616A
; Patent No. 6395278
; GENERAL INFORMATION:
; APPLICANT: Dillon, Davin C.
; APPLICANT: Harlocker, Susan Louise
; APPLICANT: Jiang, Yuqi
; APPLICANT: Xu, Jiangchun
; APPLICANT: Mitcham, Jennifer Lynn
; TITLE OF INVENTION: COMPOUNDS FOR IMMUNOTHERAPY AND DIAGNOSIS
; FILE REFERENCE: 210121.427C8
; CURRENT APPLICATION NUMBER: US/09/352,616A
; CURRENT FILING DATE: 1998-07-13
; NUMBER OF SEQ ID NOS: 472
; SOFTWARE: FASTSEQ for Windows Version 3.0
; SEQ ID NO 271
; LENGTH: 301
; TYPE: DNA
; ORGANISM: Homo sapien
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)...(301)
; OTHER INFORMATION: n = A,T,C or G
US-09-352-616A-271

Alignment Scores:
Pred. No.: 1,23e-12 Length: 301
Score: 194.50 Matches: 48
Percent Similarity: 61.39% Conservative: 14
Best Local Similarity: 47.52% Mismatches: 17
Query Match: 5.03% Indels: 23
DB: 4 Gaps: 3

US-10-045-815-4 (1-736) x US-09-352-616A-271 (1-301)

QY 637 ValThrAlaGluValGln-ProGlnCysGlyArgAlaValGlyPheSerSerGlyThrGl 656
Db 301 GTGACTGCTTCTATAAAACCCAAATGTGGCGCATGATGATTN-TCATCTGGAGGAGA 243
QY 656 uAsnProHisGlyValLysAlaValThrArgGlyGlnArgCysAlaIleAlaLeuTpph 676
Db 242 GAACCNVTCATGGGTGTAAGGAGTCCACCAAGGAAGAGGTTGCTGTGGCTCTGTGGTT 183
QY 676 eThrLeuAspProArgHisSerGluArgAspArgValGlnAlaAspLeuValLysMe 696
Db 182 CACCTTGGACCCACTTTATAGAGAAATGGAGCAATACAGGCTGATGAAGTGAATGCAAT 123
QY 696 tLeuPheSerProGluGluMetAspLeuSerGlnGluGlnProLeuAspAlaGlnGlnGl 716
Db 122 T-----CTGGATCAAGAA-----CAGCAAGG 102
QY 716 yProProGluProAlaGlnGluSerLeuSerGlySerGluSerLysProLysAspGluLe 736
Db 101 GAAGCATGAA-----CTGAATATCAACCCCTAAAGATGAGCT 66
QY 736 u 736
Db 65 A 65

RESULT 3
US-09-232-149A-271/c
; Sequence 271, Application US/09232149A
; Patent No. 6465611
; GENERAL INFORMATION:
; APPLICANT: Xu, Jiangchun
; APPLICANT: Dillon, Davin C.
; APPLICANT: Mitcham, Jennifer Lynn
; TITLE OF INVENTION: COMPOUNDS FOR IMMUNOTHERAPY OF PROSTATE
; FILE REFERENCE: 210121.427C6
; CURRENT APPLICATION NUMBER: US/09/232,149A
; CURRENT FILING DATE: 1999-01-15
; NUMBER OF SEQ ID NOS: 338
; SOFTWARE: FASTSEQ for Windows Version 3.0
; SEQ ID NO 271
; LENGTH: 301
; TYPE: DNA
; ORGANISM: Homo sapien
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)...(301)
; OTHER INFORMATION: n = A,T,C or G
US-09-232-149A-271

Alignment Scores:
Pred. No.: 1,23e-12 Length: 301
Score: 194.50 Matches: 48
Percent Similarity: 61.39% Conservative: 14
Best Local Similarity: 47.52% Mismatches: 17
Query Match: 5.03% Indels: 23
DB: 4 Gaps: 3

US-10-045-815-4 (1-736) x US-09-232-149A-271 (1-301)

QY 637 ValThrAlaGluValGln-ProGlnCysGlyArgAlaValGlyPheSerSerGlyThrGl 656
Db 301 GTGACTGCTTCTATAAAACCCAAATGTGGCGCATGATGATTN-TCATCTGGAGGAGA 243
QY 656 uAsnProHisGlyValLysAlaValThrArgGlyGlnArgCysAlaIleAlaLeuTpph 676
Db 242 GAACCNVTCATGGGTGTAAGGAGTCCACCAAGGAAGAGGTTGCTGTGGCTCTGTGGTT 183
QY 676 eThrLeuAspProArgHisSerGluArgAspArgValGlnAlaAspLeuValLysMe 696
Db 182 CACCTTGGACCCACTTTATAGAGAAATGGAGCAATACAGGCTGATGAAGTGAATGCAAT 123
QY 696 tLeuPheSerProGluGluMetAspLeuSerGlnGluGlnProLeuAspAlaGlnGlnGl 716
Db 122 T-----CTGGATCAAGAA-----CAGCAAGG 102
QY 716 yProProGluProAlaGlnGluSerLeuSerGlySerGluSerLysProLysAspGluLe 736
Db 101 GAAGCATGAA-----CTGAATATCAACCCCTAAAGATGAGCT 66
QY 736 u 736
Db 65 A 65
```



; APPLICANT: Henderson, Robert A.  
 ; APPLICANT: Kalos, Michael D.  
 ; APPLICANT: Panger, Gary R.  
 ; APPLICANT: Retter, Marc W.  
 ; APPLICANT: Stolk, John A.  
 ; APPLICANT: Day, Craig H.  
 ; APPLICANT: Vedwick, Thomas S.  
 ; APPLICANT: Carter, Darick  
 ; APPLICANT: Li, Samuel  
 ; APPLICANT: Wang, Aijun  
 ; APPLICANT: Skeky, Yasir A.W.  
 ; APPLICANT: Hepler, William  
 ; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY AND  
 ; FILE OF INVENTION: DIAGNOSIS OF PROSTATE CANCER  
 ; FILE REFERENCE: 210121.42717C17  
 ; CURRENT APPLICATION NUMBER: US/09/636,215  
 ; CURRENT FILING DATE: 2000-08-10  
 ; NUMBER OF SEQ ID NOS: 852  
 ; SOFTWARE: FastSeq for Windows Version 3.0  
 ; SEQ ID NO 271  
 ; LENGTH: 301  
 ; TYPE: DNA  
 ; ORGANISM: Homo sapien  
 ; FEATURE:  
 ; NAME/KEY: misc feature  
 ; LOCATION: (1)-(301)  
 ; OTHER INFORMATION: n = A,T,C or G  
 ; US-09-636-215-271

Alignment Scores:  
 Pred. No.: 1,23e-12 Length: 301  
 Score: 194.50 Matches: 48  
 Percent Similarity: 61.39% Conservative: 14  
 Best Local Similarity: 47.52% Mismatches: 17  
 Query Match: 5.03% Indels: 23  
 DB: 4 Gaps: 3

US-10-045-815-4 (1-736) x US-09-636-215-271 (1-301)  
 QY 637 ValThrAlaGluValGln-ProGlnCysGlyArgAlaValGlyPheSerGlyThrGl 656  
 Db 301 GTGACTGCTTCTATAAACCCAAATGTGGCGCATGATGATTN-TCACTCGAGGAGA 243  
 QY 656 uAsnProHisGlyValLysAlaValThrArgGlyCysAlaLeuAlaLeuTrpPh 676  
 Db 242 GAACNCTCATGGGTGAAGGAGTCAACCAAGGAAGAGGTGTCTGTGGTCTGTGGTT 183  
 QY 676 eThrLeuAspProArgHisSerGluArgAspArgValGlnAlaAspLeuValLysMe 696  
 Db 182 CACCTTGGACCCACTTTATAGAGATTGGAGCGAATACAGCTGATGAGTGCAT 123  
 QY 696 tLeuPheSerProGluGluMetAspLeuSerGlnGluGlnProLeuAspAlaGlnGln 716  
 Db 122 T-----CTGGATCAAGAA-----CAGCAAGG 102  
 QY 716 yProProGluProAlaGlnGluSerLeuSerGlySerGluSerLysProLysAspGluLe 736  
 Db 101 GAAGCATGAA-----CTGAAATATCAACCTTAAGATGAGCT 66  
 QY 736 u 736  
 Db 65 A 65

RESULT 6  
 US-09-685-166A-271/c  
 ; Sequence 271, Application US/09685166A  
 ; Patent No. 6630305  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Xu, Jiangchun  
 ; APPLICANT: Dillon, Davin C.  
 ; APPLICANT: Mitcham, Jennifer L.  
 ; APPLICANT: Harlocker, Susan L.  
 ; APPLICANT: Jiang, Yuqi

QY 736 u 736  
 Db 65 A 65

RESULT 4  
 US-09-159-812-271/c  
 ; Sequence 271, Application US/09159812A  
 ; Patent No. 6613872  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Xu, Jiangchun  
 ; APPLICANT: Dillon, Davin C.  
 ; TITLE OF INVENTION: COMPOUNDS FOR IMMUNOTHERAPY OF  
 ; FILE OF INVENTION: PROSTATE CANCER AND METHODS FOR THEIR USE  
 ; FILE REFERENCE: 210121.428C5  
 ; CURRENT APPLICATION NUMBER: US/09/159,812A  
 ; CURRENT FILING DATE: 1998-09-23  
 ; NUMBER OF SEQ ID NOS: 306  
 ; SOFTWARE: FastSeq for Windows Version 3.0  
 ; SEQ ID NO 271  
 ; LENGTH: 301  
 ; TYPE: DNA  
 ; ORGANISM: Homo sapien  
 ; FEATURE:  
 ; NAME/KEY: misc feature  
 ; LOCATION: (1)-(301)  
 ; OTHER INFORMATION: n = A,T,C or G  
 ; US-09-159-812-271

Alignment Scores:  
 Pred. No.: 1,23e-12 Length: 301  
 Score: 194.50 Matches: 48  
 Percent Similarity: 61.39% Conservative: 14  
 Best Local Similarity: 47.52% Mismatches: 17  
 Query Match: 5.03% Indels: 23  
 DB: 4 Gaps: 3

US-10-045-815-4 (1-736) x US-09-159-812-271 (1-301)  
 QY 637 ValThrAlaGluValGln-ProGlnCysGlyArgAlaValGlyPheSerGlyThrGl 656  
 Db 301 GTGACTGCTTCTATAAACCCAAATGTGGCGCATGATGATTN-TCACTCGAGGAGA 243  
 QY 656 uAsnProHisGlyValLysAlaValThrArgGlyCysAlaLeuAlaLeuTrpPh 676  
 Db 242 GAACNCTCATGGGTGAAGGAGTCAACCAAGGAAGAGGTGTCTGTGGTCTGTGGTT 183  
 QY 676 eThrLeuAspProArgHisSerGluArgAspArgValGlnAlaAspLeuValLysMe 696  
 Db 182 CACCTTGGACCCACTTTATAGAGATTGGAGCGAATACAGCTGATGAGTGCAT 123  
 QY 696 tLeuPheSerProGluGluMetAspLeuSerGlnGluGlnProLeuAspAlaGlnGln 716  
 Db 122 T-----CTGGATCAAGAA-----CAGCAAGG 102  
 QY 716 yProProGluProAlaGlnGluSerLeuSerGlySerGluSerLysProLysAspGluLe 736  
 Db 101 GAAGCATGAA-----CTGAAATATCAACCTTAAGATGAGCT 66  
 QY 736 u 736  
 Db 65 A 65

RESULT 5  
 US-09-636-215-271/c  
 ; Sequence 271, Application US/09636215  
 ; Patent No. 6620922  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Xu, Jiangchun  
 ; APPLICANT: Dillon, Davin C.  
 ; APPLICANT: Mitcham, Jennifer L.  
 ; APPLICANT: Harlocker, Susan L.  
 ; APPLICANT: Jiang, Yuqi





```

/ OTHER INFORMATION: n=a or c or g or t
/ NAME/KEY: misc feature
/ LOCATION: (765001)..(780000)
/ OTHER INFORMATION: n=a or c or g or t
/ NAME/KEY: misc feature
/ LOCATION: (780001)..(795000)
/ OTHER INFORMATION: n=a or c or g or t
/ NAME/KEY: misc feature
/ LOCATION: (795001)..(810000)
/ OTHER INFORMATION: n=a or c or g or t
/ NAME/KEY: misc feature
/ LOCATION: (810001)..(825000)
/ OTHER INFORMATION: n=a or c or g or t
/ NAME/KEY: misc feature
/ LOCATION: (825001)..(840000)
/ OTHER INFORMATION: n=a or c or g or t
/ NAME/KEY: misc feature
/ LOCATION: (840001)..(855000)
/ OTHER INFORMATION: n=a or c or g or t
/ NAME/KEY: misc feature
/ LOCATION: (855001)..(870000)
/ OTHER INFORMATION: n=a or c or g or t
/ NAME/KEY: misc feature
/ LOCATION: (870001)..(885000)
/ OTHER INFORMATION: n=a or c or g or t
/ NAME/KEY: misc feature
/ LOCATION: (885001)..(900000)
/ OTHER INFORMATION: n=a or c or g or t
/ NAME/KEY: misc feature
/ LOCATION: (900001)..(915000)
/ OTHER INFORMATION: n=a or c or g or t
/ NAME/KEY: misc feature

```

## Alignment Scores:

```

Pred. No.: 0.00584 Length: 1230025
Score: 162.50 Matches: 146
Percent Similarity: 33.53% Conservative: 86
Best Local Similarity: 21.10% Mismatches: 257
Query Match: 4.20% Indels: 203
DB: 4 Gaps: 30

```

US-10-045-815-4 (1-736) x US-09-198-452A-1 (1-1230025)

```

Qy 51 TrpProGlyValValLeuSerMetGluArgAlaLeuArgSerArgAlaAlaLeuArgAla 70
Db 1082828 TGGCCCAAGTG-----GAAGTATCCGACGAGGCGATACCTCTTCA 1082784
Qy 71 LeuA:GLeu-----ArgCysArgThrGlnCysAla 80
Db 1082783 CAACGACAAATTGATTTTGGCCCTTATGTAAGTGCTGAAGCCTGCCACACTCTCTTA--- 1082727
Qy 81 AlaAspPheProTribLeuAspProAspTribSerProSerProAlaGlnAlaSerGly 100
Db 1082726 -----GAGTTATCAGCAATGTTTCTCT----- 1082703
Qy 101 AlaGlyAlaLeuArgAspLeuSerPheGlyGlyLeuLeuArgArgAlaAlaCys--- 119
Db 1082702 -----CTGAGGACTTGTTCAGATCGAGAGTTTGCCTTAAGGAAACGCCCTTGCATT 1082652
Qy 120 -----LeuArgCysLeuGlyProProAlaAlaHisSerLeuSerGluGluMet 136
Db 1082651 CTTTATGACATGAAGCGTGCTTCTCTCTGTGTGCGCTATTATGCTCCTCGAAGAATAT 1082592
Qy 137 GluLeuGluPheArgGlyArgSerProTyrAsnTyrLeuGlnValAlaTyrPheGlyAla 156
Db 1082591 CAAGCAACCTTAGACAAG-----GCGATCCTCTTCTTTAAA 1082556
Qy 157 AsnLysLeuGluLysAlaValAlaAlaAlaHisThrPheValGlyAsnProGluHis 176
Db 1082555 GGAAAGATCGAAGAGTGGTCAAGGACTTAGAAAAGTGTATCAAAAGGCTTCGCGATAAC 1082496
Qy 177 MetGluMetGlnAsnLeuAspTyrTyrGlnThrMetSerGlyValLysGluAlaAsp 196

```

```

Db 1082495 TTAGAATTTGAGCAAGCTGCGAATTACTACCGACAGTGTCTCCCTAATTAAGCAAGCTATG 1082436
Qy 197 PheLysAspLeuGluThrGlnProHisMetGlnGluPhe----- 209
Db 1082435 GCAAGCAACAGGTGGAAAAATTTCAATTTCAAAATATTGATGCTCTTGGTCTTTACAGA 1082376
Qy 210 -----ArgLeuGlyVal 213
Db 1082375 CATAAAACACGACCATCTCCTACGCTACTACTGTCCGCTCAGGGAAGCTTCTTGGGCA 1082316
Qy 214 ArgLeuTyrSer---GluGluGlnProGlnGluAlaValProHisLeuGluAlaLeu 232
Db 1082315 CGCAATTTCTTTCTTTGAGAATGCTCAAGAAGATCAAGACTTGTCTCTCTTTTATT 1082256
Qy 233 GlnGluTyrPheValAla-----TyrGluGluCysArgAlaLeuCysGluGlyProTyr 250
Db 1082255 TTGCAATATTATGTTAGCAACCTACATTCTTAAAGAAATCTGCACACCTCTTCCCTA 1082196
Qy 251 AspTyrAspGlyTyrAsnTyrLeuGluTyrAsnAlaAsp-----LeuPheGln 266
Db 1082195 GAAATTCCTACTCTCTCTACGTA---TTGAATGCGGAGTCTCTCCCGCCTACGTTCT 1082139
Qy 267 AlaIleThrAspHisTyrIleGlnValLeu-----AsnCysLysGlnAsn 281
Db 1082138 CCAAAACAGGATATGAAAAGAACTTTTGATCTAGCCTATCGCAATGCCAAGCGGTAT 1082079
Qy 282 CysValThrGluLeuAlaSerHisProSerArgGluLysProPheGluAspPheLeu--- 300
Db 1082078 GCAGCAACTACACTTCTCTCATCAAGCTCCCTACCAAGACTTTAAGATAATACTCGCG 1082019
Qy 301 ProSerHisTyrAsnTyrLeuGlnPheAlaTyrTyrAsnIleGlyAsnTyrThrGlnAla 320
Db 1082018 ATGTCATATCCGTATCGGATAGTATCGATAGTCTCTAGTATAGTCT-----CATATGCAAGT 1081965
Qy 321 GlyGluCysAlaLysThrTyrLeuLeuPhePheProAsnAspGluValMetAsnGlnAsn 340
Db 1081964 GCTCAGGCTACTGGAGTGACATTTGTTATTTGAAAATAATAGGGTTCGATCCCAAGCAA--- 1081908
Qy 341 LeuAlaTyrTyrAlaAlaMetLeuGluGluHisThrArgSerIleGlyProArgGlu 360
Db 1081907 -----TATCGTACCTTTTCCATA-----GAT 1081887
Qy 361 SerAlaLysGluTyrArgGlnArgSerLeuLeuGluLysGluLeuLeuPhePheAlaTyr 380
Db 1081886 TCTGAAAGACTCAGAAATGACCTGGCTCTATTAGAAGAAGTCTCTTA----- 1081839
Qy 381 AspValPheGlyIleProPheValAspProAspSerTribThrProGluGluValIlePro 400
Db 1081838 -----CGAAGGTTTTCATTCACGACACAGCTCTTCTCGATGATGATGTTG--- 1081794
Qy 401 LysArgLeuGlnGluLysGlnLysSerGluArgGluThrAlaValArgIleSerGlnGlu 420
Db 1081793 -----GTTGATGGGGGAAAGCCACTATAACAGACAAAAAATAATCCAACT 1081743
Qy 421 IleGlyAsnLeuMetLysGluIleGluThrLeuValGluGluLysThrLysGluSer--- 439
Db 1081742 CTCAATCTTACAGGAATTCAGTTGTTACTATCGCAAGAAAAAGTAATCATAGTCGA 1081683
Qy 440 ---LeuAspValSerArgLeuThrArgGluGlyGlyProLeuLeuTyrGluGlyIleSer 458
Db 1081682 GGGTTAAATAGGAGNAGATTTCTCGAGAGCTTCCCA-----GAGGTTTTTCT 1081632
Qy 459 LeuThrMetAsnSerLysLeuLeuAsnGlyTyrGlnArgValValMetAspGlyValIle 478
Db 1081631 CTTCCCGGACTTCTAAATTTATTACAAATTTTTCAG-----ATCCTT 1081590
Qy 479 SerAspHisGluCysGlnGluLeuGlnArgLeuThrAsnValAlaAlaThrSerGlyAsp 498
Db 1081589 CGTGAT-----GAG 1081581
Qy 499 GlyTyrArgGlyGlnThrSerProHisThrProAsnGluLysPheTyrGlyValThrVal 518
Db 1081580 GCACATCGCTTTGCGATTAGTAGCAC-----AGGAAAAAAGAGAAAGGCTTTA 1081530

```

Qy	519	PhelysAlaLeuLysLeuGlyGlnGluGlyValProLeuGlnSerAlaHisLeuTyr	538	Percent Similarity: 34.89%	Conservative: 54		
Db	1081529	TTTGAACAGGAAAAATCCCTCGGAATCGGAGAGTA	1081494	Best Local Similarity: 22.25%	Mismatches: 160		
				Query Match: 3.17%	Indels: 118		
				DB: 4	Gaps: 20		
Qy	539	TyrAsnValThrGluLysValArgIleMetGluSerTyrPheArgLeuAspThrPro	558	US-10-045-815-4 (1-736) x US-09-738-946-5 (1-2595)			
Db	1081493	-----AAGCGAAAAAGATTCTTCAGAAA-----	1081470	Qy	355	SerIleGlyProArgGluSerAlaLysGluTyrArgGlnArgSerLeuLeuGluLys-Gl	374
Qy	559	LeuTyrPheSerTyrSerHisLeuValCysArgThrAlaIleGluGluValGlnAlaGlu	578	Db	782	AGCTGTGGCATCTGCCAATCTGGAATGAGGGGCGCGCGCAACTGCTCCACGCCCA	841
Db	1081469	-----TTTAAAGTTGGAACAGGTATATCTAGTCCCGAGGAGAGCTGGAAGCT---	1081419	Qy	374	uLeuLeuPhePheAlaTyrAspValPheGlyIleProPheValAspProAspSerTyrTh	394
Qy	579	ArgLysAspAspSerHisProValHisValAspAsnCysIleLeuAsnAlaGluThrLeu	598	Db	842	GATGATG-----ACGCCTGCCACCCAGGGCGAGAGTTGG--	875
Db	1081419	-----	1081419	Qy	394	rProGluGluVal-----IleProLysArgLeuGlnGluLys	406
Qy	599	ValCysValLysGluProProAlaTyrThrPheArgAspTyrSerAlaIleLeuTyrLeu	618	Db	876	-----CCCCCGGGGTAGACAACCTGCTGAATCTCTCGGCGAGCGCGCAGGAGAAGGC	934
Db	1081418	-----ATTCTGGGCTAACTAAGAAAGATATTGCAGTCTTTTGCAGG	1081374	Qy	406	sGlnLysSerGluArgGluThrAlaValArgIleSerGln-----	419
Qy	619	AsnGlyAspPheAspGlyGlyAsn---PheTyrPheThrGluLeuAspAlaLysThrVal	637	Db	935	TCAGCTCGGGAACGGAACAGGTCAAGAGCAGCAGCAGCAGCATCATCACCATAA	994
Db	1081373	CAAAAAGACTTCAACAAATCTGATTAAGTCTACCTTCAGATTCCTCAAGTAAGAAGTCT	1081314	Qy	420	-----GluIleGlyAsnLeuMetLysGluIleGluThrLeuVa	432
Qy	638	ThrAlaGluValGlnProGlnCysGlyArgAlaValGlyPheSerSerGlyThrGluAsn	657	Db	995	TGGCGAAAAGAGCTCCAGCTATCAATCGGTCTGGCGGATGCCAGCTTCATGGGATCAGG	1054
Db	1081313	TCAGGTGAA-----GGAAGA-----TCTTGAACAGAGACT	1081284	Qy	432	IcGluGluLysThrLysGluSerLeu-----AspValSerArgLeuThrArgGluGl	449
Qy	658	ProHisGlyValLysAlaValThrArgGlyGlnArgCysAlaIleAlaLeuTrpPheThr	677	Db	1055	AAAGTGGCGCCGCTATGAGGATCTGTCCGTAAACATCATCAGCAGCATGAACCATATGG	1114
Db	1081283	TCCTCTCAATTTCTCATCAATACGATCTTTTGGTTGTGTCAACCTTTAATGATCAGG	1081224	Qy	449	VGlyProLeuLeuTyrGluGlyIleSerLeuThrMetAsnSerLysLeuLeuAsnGlyTy	469
Qy	678	Leu-----AspProArgHisSerGlu	684	Db	1115	TTGTCCGTGGTGGACGACTTCTCTGGGATGGAGACGGGCGCTGAAGATCCTCAACGAGGT	1174
Db	1081223	TGGACAGCGGCCCATGAAGTCTCTGTATATCTGAA	1081188	Qy	469	rGlnArgValValMetAspGlyValIleSerAspHisGluCysGlnGluLeuGlnArgLe	489
RESULT 9				Db	1175	TGCAAGCATGTACAACGAGGAGCCCTCCAGATGCG-----CAAGTGGT	1219
US-09-738-946-5				Qy	489	uThrAsn-----ValAlaAlaThrSerGlyAspGlyTyrArgGlyGln-----	503
; Sequence 5, Application US/09738946				Db	1220	GACCAACAGAGCGCCGATGACCCGCGTGGCGGTGACAGATCCGAGGCGATAAGAT	1279
; Patent No. 6579701				Qy	504	-----ThrSerProHisThrProAsnGluLysPhe-----Ty	514
; GENERAL INFORMATION:				Db	1280	CAAGTGGGTGGTGGCAATGAGCCGGCTGCAGCAATGTCTGTGTATCTGACCAATCAGAT	1339
; APPLICANT: EXELIXIS, INC.				Qy	514	rGlyValThrValPheLysAlaLeuLysLeuGlyGlnGluGlyLysValProLeuGlnSe	534
; TITLE OF INVENTION: DROSOPHILA HOMOLOGUES OF GENES AND PROTEINS IMPLICATED IN CANCER				Db	1340	TGACTCTGTGTGTATCGTTTAAACAGCATGAAGATAATGGCATCTTTGGGC-----	1391
; TITLE OF INVENTION: METHODS OF USE				Qy	534	rAlaHisLeuTyrTyrAsnValThrGluLysValArgArgIleMetGluSerTyrPheAr	554
; FILE REFERENCE: EX00-043C				Db	1392	-----AACTACCACATCAGGAGCGCAGGAGGCAATGGTCTGTGTAT-----	1436
; CURRENT APPLICATION NUMBER: US/09/738,946				Qy	554	gLeuAspThrProLeuTyrPheSerTyrSerHisLeuValCysArgThrAlaIleGluGl	574
; CURRENT FILING DATE: 2000-12-14				Db	1437	-----CCGGATCGGGAACCTCACTACGTC-----	1460
; PRIOR APPLICATION NUMBER: 60/170,832				Qy	574	uValGlnAlaGluArgLysAspSerHisProValHisValAspAsnCysIleLeuAs	594
; PRIOR FILING DATE: 1999-12-14				Db	1461	-----ATGCATGTGGCAAT-----	1475
; PRIOR APPLICATION NUMBER: 60/170,838				Qy	594	nAlaGluThrLeuValCysValLysGluProProAlaTyrThrPheArgAspTyrSerAl	614
; PRIOR FILING DATE: 1999-12-14				Db	1476	-----CCCCAAAGAGATGGCGCGGTATAACGGC	1504
; PRIOR APPLICATION NUMBER: 60/178,580				Qy	614	alleLeuTyrLeuAsnGlyAspPheAsp-----GlyGlyAsnPheTyrPhen	630
; PRIOR FILING DATE: 2000-01-28				Db	1505	CATATACTACCTGAATATCACTGGGATGCGGGAGAGTGGCGGATTTCTCGGAATTCG	1564
; PRIOR APPLICATION NUMBER: 60/185,879				Qy	630	rGluLeuAspAlaLysThrValThrAlaGluValGlnProGlnCysGlyArgAlaValGl	650
; PRIOR FILING DATE: 2000-02-29							
; PRIOR APPLICATION NUMBER: 60/185,880							
; PRIOR FILING DATE: 2000-02-29							
; PRIOR APPLICATION NUMBER: 60/186,150							
; PRIOR FILING DATE: 2000-03-01							
; PRIOR APPLICATION NUMBER: 60/189,701							
; PRIOR FILING DATE: 2000-03-15							
; NUMBER OF SEQ ID NOS: 14							
; SOFTWARE: PatentIn version 3.0							
; SEQ ID NO 5							
; LENGTH: 2595							
; TYPE: DNA							
; ORGANISM: Drosophila melanogaster							
US-09-738-946-5							
Alignment Scores:							
Pred. No.:							
Score:							

0.00492 Length: 2595  
122.50 Matches: 95



```
; LENGTH: 1662
; TYPE: DNA
; ORGANISM: Enterococcus faecalis
US-09-134-000C-666

Alignment Scores:
Pred. No.: 0.0122 Length: 1662
Score: 116.00 Matches: 73
Percent Similarity: 35.62% Conservative: 67
Best Local Similarity: 18.58% Mismatches: 119
Query Match: 3.00% Indels: 134
DB: 4 Gaps: 20

US-10-045-815-4 (1-736) x US-09-134-000C-666 (1-1662)

QY 169 PhePheValGlyAsnProGlu-----HisMetGluMetGlnGlnAsnLeuAspTyr 185
DB 622 TTTTATACAGAAACCTGAGATCTTTATCATGAATTAACAAAGCCACCCAGAT 681
QY 186 TyrGlnThrMet-----SerGlyValLysGluAlaAspPhe 197
DB 682 TATCAAGTGATGATCAACTGGAAAGCCGTTGCAAGCCGTATTGTCAAAGAAGATGTT 741
QY 198 -----LysAspLeuGluThrGlnProHisMetGlnGluPheArg 210
DB 742 CGTTTGTGTTACGATTACCTAAAGATGCTTTCGTCAACCGTACATGGCCAAATAC--- 798
QY 211 LeuGlyValArgLeuTyrSerGluGluGlnProGlnGluAlaValProHisLeuGluAla 230
DB 799 TTAACGGTGAATATCTGTCAAAGAATG-----828
QY 231 AlaLeuGlnGluTyrPheValAlaTyrGluGluCysArgAlaLeuCysGluGlyProTyr 250
DB 829 -----837
QY 251 AspTyrAspGlyTyrAsnTyrLeuGluTyrAsnAlaAspLeuPheGlnAla---267
DB 838 TCTTGGATAGCTTTCATTGATTCAGGAGAAACGCTTTTGAAGGAAGTCTTTTG 897
QY 268 -----IleThrAspHisTyrIleGlnValLeuAsnCysLysGln 280
DB 898 GCTCAACGACCAACAAATGAATGAAACCAATTTATTCAAGTG---AAAATTGAAAC 954
QY 281 AsnCysValThrGluLeuAlaSerHisProSerArgGluLysProPheGluAspPheLeu 300
DB 955 AATGGTAGCTTAACCATTCGTCGACAAAACACGAGAA-----993
QY 301 ProSerHisTyrAsnTyrLeuGlnPheAlaTyrTyrAsnIleGlyAsnTyrThrGlnAla 320
DB 994 -----ACATTTTCAAAATTTGTTAACATTTGAAGATACT 1026
QY 321 GlyGluCysAlaLysThrTyrLeuLeuPhePheProAsnAspGlu-----ValMetAsn 338
DB 1027 GGCACATTTGGGAATGAATATATCTTTTCAAAACCGACCGAAGACCAAGGAATTACAAAC 1086
QY 339 GlnAsnLeuAlaTyrTyrAlaAlaMetLeuGlyGluGluHisThrArgSerIleGlyPro 358
DB 1087 GAATATGTAACG-----GCTGAATCAGCAATTAAGAAATTC-----CCT 1128
QY 359 ArgGluSerAlaLysGluTyrArgGlnArgSerLeuLeuGluLysGluLeuPhePhe 378
DB 1129 GTAAAGACAGATTATCAATTAATAACAAACCGTCATGTTACCA-----1170
QY 379 AlaTyrAspValPheGlyIleProPheValAspProAspSerTyrThrProGluGluVal 398
DB 1171 -----GTTGCGGCAGAT-----1182
QY 399 IleProLysArgLeuGlnGluLysGlnLysSerGluArgGluThrAlaValArgIleSer 418
DB 1183 -----GAACGATTAGAAGAAGAACAAAGAGCTTCTCGTGAGTTTAGAAGAACGCTTGCA 1236
QY 419 GlnGluIleGlyAsnLeuMetLys---GluIleGluThrLeuValGluGluLysThrLys 437
```

```
Db 1237 CAACGTTCAACGACGCTTCGCCGTTTCGAAATAACGACAATGGTT-----1281
QY 438 GluSerLeuAspValSerArgLeuThrArgGluGlyGlyProLeuLeuTyrGluGlyIle 457
DB 1282 -----ACAAATGATCAAGAAAGAAAGTAACCAATTTGTTTGA-----1317
QY 458 SerLeuThrMetAsnSerLysLeuLeuAsnGlyTyrGlnArgValValMet---AspGly 476
DB 1318 ---ACAACCATTAATAATCAAAATCAAGATCATCGTTTACGGGTGCTTTTCCCAACTGGC 1374
QY 477 ValIleSerAsp---HisGluCysGlnGluLeuGlnArgLeuThrAsnValAlaThr 495
DB 1375 ATGGTCACAGAAACACACGAGCGGATATATTATGAAGTCGTGACACGACCAATCAA 1434
QY 496 SerGlyAspGlyTyrArgGlyGlnThrSerPro-----506
DB 1435 GTAAGCGACATCGGAAATCCAAACACCGCAGCACCACCAAGCGTTTGTGAATGTT 1494
QY 507 HisThrProAsnGluLysPheTyrGlyValThrValPhe 519
DB 1495 CATGACCAAAACAAA-----GGCGTGACTATCTTC 1524

RESULT 12
US-09-604-694B-1
; Sequence 1, Application US/09604694B
; Patent No. 6579522
; GENERAL INFORMATION:
; APPLICANT: BROUGH, DOUGLAS E
; APPLICANT: KING, C R
; APPLICANT: KOVESDI, IMRE
; APPLICANT: SCHALBLE, JASPER J
; TITLE OF INVENTION: REPLICATION DEFICIENT ADENOVIRAL TNF VECTOR
; FILE REFERENCE: 202028
; CURRENT APPLICATION NUMBER: US/09/604,694B
; CURRENT FILING DATE: 2003-01-10
; NUMBER OF SEQ ID NOS: 2
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 1
; LENGTH: 32798
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Synthetic construct
US-09-604-694B-1

Alignment Scores:
Pred. No.: 2.17 Length: 32798
Score: 115.50 Matches: 96
Percent Similarity: 30.94% Conservative: 42
Best Local Similarity: 21.52% Mismatches: 146
Query Match: 2.98% Indels: 162
DB: 4 Gaps: 20

US-10-045-815-4 (1-736) x US-09-604-694B-1 (1-32798)

QY 42 GlyThrAlaAlaTyrAlaArgGlyAspTyrProGlyValValLeuSerMetGluArgAla 61
DB 24125 GGAGATGCAAGCCAGTAGAGAAAC-----ATTGCCTACACCTTCGACAGG 24175
QY 62 LeuArgSer-----ArgAlaAlaLeuArgAlaLeuArgLeuArg 74
DB 24176 CTACGTACCCAGCGCTGCAAGATCTCCAACTGGAGCTCTGCAACCTGCTCTCCTA--- 24232
QY 75 CysArgThrGlnCysAlaAlaAspPheProTyrGluLeu-----AspProAspTyrSer 92
DB 24233 -----CCTTGGAAATTTGCACGAAACCCGCTTGGGCA 24265
QY 93 ProSerPro-----AlaGlnAlaSerGlyAlaGlyAlaLeuArgAsp-----106
DB 24266 AAACGTGCTTCATTCCACGCTCAAGGGGAGCGCGCGAGCTACGTCGCGAGCTGCGT 24325
QY 107 -----LeuSerPheGlyLeuLeuArgArgAlaAla-CysLeuArg 121
```



Db 24326 TTACTATTCTATGCTACACCTGGCAGACGCGCATGGCGCTTTGGCAGCAGTCTTGA 24385  
Qy 121 GARGCysLeuGlyProProAlaAlaHisSerLeuSerGluGluMetGluLeu----- 138  
Db 24386 GGAGTGC-----AACCTCAAGGAGCTGCAGAACTGCTTAAGCA 24424  
Qy 139 -----GluPheArgLysArgSerProTyrAsnTyrLeuGlnVa 151  
Db 24425 AAACCTTGAGGACCTATGAGCGGCTTCAACGAGCGCTCC----- 24464  
Qy 151 LAlaTyrPheLysIleAsnLysLeuGluLysAlaValAlaAlaHisThrPheVa 171  
Db 24465 -----GTGCCCGGCACCTGGCGGACAT 24487  
Qy 171 lGlyAsnProGluHisMet-----GluMetGlnGlnAsnLeuAspTyrTyrGlnThrMe 189  
Db 24488 CATTTTCCCGAACGCTGCTTAAACCCCTGCAACAGGGTCTG----- 24530  
Qy 189 tSeiclyValLysGluAlaAspPheLysAspLeuGluThrGlnProHisMetGlnGluPh 209  
Db 24531 -----CCAGATTCCACGTCACAGCATGTTCGAACTT 24565  
Qy 209 eArgLeuGlyValArgLeuTyrSerGluGluGlnProGlnAlaValProHisLeuGl 229  
Db 24566 TAGGAACCTTATCTCA-----GAGCGCTCAGGAATCTTGCCCGCCACCTG 24610  
Qy 229 uAlaAlaLeuGlnGluTyrPheVal-----AlaTyrGluGluCysArgAlaLeuCysGl 247  
Db 24611 CTGTGCACCTTCTAGCGACTTGTGCGCAATTAAGTACCGCAATTCGCGCGCTTGG 24670  
Qy 247 uGlyPro--TyrAspTyrAspGlyTyrAsnTyrLeuGluTyrAsnAlaAspLeuPheGl 266  
Db 24671 GGGCCACTGCTACCTTCGAGTAGCACTACCTTGCCTACCACTGCATCAATGGA 24730  
Qy 266 rAlaIleThrAspHisTyrIleGlnValLeuAsnCysLysGlnAsnCysValThr----- 284  
Db 24731 AGACGTGAGCGGTGAGCGTCTACTGGAGTGTCACTGTGCTGCAACCTATGCACCCGCA 24790  
Qy 285 -----G1 285  
Db 24791 CGGCTCCCTGGTTTGCATTCGACGCTCTTAACGAAAGTCAATATTCGTACTTTGA 24850  
Qy 285 uLeuAlaSerHisProSerArgGluLysPro----- 295  
Db 24851 GCTGACGGGTCCCTCGCCTGACGAAAGTCCGCGCTCCGGGTGAACTCACTCCGGG 24910  
Qy 296 -----PheGluAspPheLeuProSerHisTyrAsnTyrLeuGlnPh 309  
Db 24911 GCTGTGGACGTCGGCTTACCTTGGCAATTTGTACCTGAGGACTACCAACCCACGAGAT 24970  
Qy 309 eAlaTyrTyrAsnIleGlyAsnTyrThrGlnAlaGlyLysAlaLysThrTyrLeuLe 329  
Db 24971 TAGGTTCTAC----- 24980  
Qy 329 uPhePheProAsnAspGluValMetAsnGlnAsnLeuAlaTyrTyrAlaAlaMetLeuGl 349  
Db 24981 -----GAAGACCAATCCCGCCCGCCCAATTCGCGGCTTACCGCTCGGTCATTAC 25030  
Qy 349 yGluGluHisThrArgSerIleGlyProArgGluSerAlaLysGluTyrArgGlnArgse 369  
Db 25031 CCAGGGCCACATT-----CTTGCCCAATTGCAAGCCATCAACAAAGCCGCGCAAGATT 25084  
Qy 369 rLeuLeuGluLysGluLeuLeuPheAlaTyrAspValPheGlyIleProPheValas 389  
Db 25085 TCTGTACGAAAGGACCG-----GGGGT---TACTTGA 25117  
Qy 389 pProAspSerTrpThrProGluGluValIleProLysArgLeuGlnGluLysGlnLysse 409  
Db 25118 CCCCAGTCCGCGAGGAGCTCAACCCATCCCCCGCCGCGGCGCTTATCAGCAGCA 25177  
Qy 409 rGluArgGluThrAla 414  
Db 25178 GCCCGGGCCCTTGT 25193

## RESULT 13

US-08-735-609-4  
; Sequence 4, Application US/08735609  
; Patent No. 5955360  
; GENERAL INFORMATION:

; APPLICANT: Chamberlain, Jeffrey S.  
; APPLICANT: Amalfitano, Andrea  
; APPLICANT: Hauser, Michael A.  
; APPLICANT: Kumar-Singh, Rajendra  
; APPLICANT: Hartigan-O'Connor, Dennis J.  
; TITLE OF INVENTION: IMPROVED ADENOVIRUS VECTORS

; NUMBER OF SEQUENCES: 15  
; CORRESPONDENCE ADDRESS:

; ADDRESSEE: Medlen & Carroll, LLP  
; STREET: 220 Montgomery Street, Suite 2200

; CITY: San Francisco

; STATE: California

; COUNTRY: United States Of America

; ZIP: 94104

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Floppy disk

; COMPUTER: IBM PC compatible

; OPERATING SYSTEM: PC-DOS/MS-DOS

; SOFTWARE: Patent In Release #1.0, Version #1.30

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/08/735,609

; FILING DATE:

; CLASSIFICATION: 435

; ATTORNEY/AGENT INFORMATION:

; NAME: Ingolia, Diane E.

; REGISTRATION NUMBER: 40,027

; REFERENCE/DOCKET NUMBER: UM-02484

; TELECOMMUNICATION INFORMATION:

; TELEPHONE: (415) 705-8410

; TELEFAX: (415) 397-8338

; INFORMATION FOR SEQ ID NO: 4:

; SEQUENCE CHARACTERISTICS:

; LENGTH: 34303 base pairs

; TYPE: nucleic acid

; STRANDEDNESS: double

; TOPOLOGY: linear

; MOLECULE TYPE: other nucleic acid

; DESCRIPTION: /desc = "DNA"

US-08-735-609-4

Alignment Scores:

Pred. No.: 2,34 Length: 34303

Score: 115.50 Matches: 96

Percent Similarity: 30.94% Conservative: 42

Best Local Similarity: 21.52% Mismatches: 146

Query Match: 2.98% Indels: 162

DB: 2 Gaps: 20

US-10-045-815-4 (1-736) x US-08-735-609-4 (1-34303)

Qy 42 GlyThrAlaAlaTyrAlaArgGlyAspTrpProGlyValValLeuSerMetGluArgAla 61  
Db 26620 GGAGATGCGACGCAAGCTAGAGGAAAC-----ATTGCACCTACACCTTTTCGACACGG 26670

Qy 62 LeuArgSer-----ArgAlaAlaLeuArgAlaLeuArgLeuArg 74  
Db 26671 CTACGTACCCAGGCGCTGCAAGATCTCCACCTGGAGCTCTGCAACCTGGTCTCTTA--- 26727

Qy 75 CysArgThrGlnCysAlaAlaAspPheProTyrGluLeu-----AspProAspTrpSer 92  
Db 26728 -----CCTTGAATTTTGCACGAAACCGCTTGGCA 26760

Qy 93 ProSerPro-----AlaGlnAlaSerGlyAlaGlyAlaLeuArgAsp----- 106  
Db 26761 AAACGTGCTTCATTCACGCTCAAGGCGAGCGCGCGGCTACCTCCGCGACTGGGT 26820

Qy 107 -----LeuSerPhePheGlyGlyLeuLeuArgAlaAla-CysLeuAr 121

Db 26821 TTACTTATTTCTATGCTACACCTGCGCAGCGCCATGGCGCTTTGGCAGCAGTGTCTTGA 26880  
QY 121 gArgCysLeuGlyProPrCAlaAlaHisSerLeuSerGluGluMetGluLeu----- 138  
Db 26881 GGAGTGC-----AACTCAAGAGCTGCGAGAAACTGCTAAAGCA 26919  
QY 139 -----GluPheArgCysArgSerProTyrAsnTyrLeuGlnVa 151  
Db 26920 AAACCTGAAGACCTATGACGGCTTCAACGAGCGCTCC----- 26959  
QY 151 lAlaTyrPheLysIleAsnLysLeuGluLysAlaValAlaAlaAlaHisThrPhePheVa 171  
Db 26960 -----GTGCCCGCGCACCTGGCGGACAT 26982  
QY 171 lGlyAsnProGluHisMet-----GluMetGlnGlnAsnLeuAspTyrTyrGlnThrMe 189  
Db 26983 CATTTTCCCGAAGCGCTGCTTAAACCCCTGCACAGGGTCTG----- 27025  
QY 189 tSerGlyValLysGluAlaAspPheLysAspLeuGluThrGlnProHisMetGlnGluPh 209  
Db 27026 -----CCAGACTTCACCATGCAAGCATGTTGCAGAACTT 27060  
QY 209 eArgLeuGlyValArgLeuTyrSerGluGluGlnProGlnGluAlaValProHisLeuG1 229  
Db 27061 TAGGAACTTTATCCTA-----GAGCGCTCAGAACTTGTGCCGCCACCTG 27105  
QY 229 uAlaAlaLeuGlnGluTyrPheVal-----AlaTyrGluGluCysArgAlaLeuCysG1 247  
Db 27106 CTGTGCACCTCTAGCGACTTGTGCCATTAAAGTACCGCAATGCCCTCGCGCTTTG 27165  
QY 247 uGlyPro-----TyrAspTyrAspGlyTyrAsnTyrLeuGluTyrAsnAlaAspLeuPheG1 266  
Db 27166 GGGCCACCTGCTACCTCTGACAGTAGCAACTGCTTGCCTACCACCTGACATAATGGA 27225  
QY 266 nAlaAlaThrAspHisTyrIleGlnValLeuAsnCysLysGlnAsnCysValThr----- 284  
Db 27226 AGACGTGAGCGGTGACGGTCTACTGGAGTGTCACCTGCTGCTGCACTATGCACCCGCA 27285  
QY 285 -----G1 285  
Db 27286 CCGCTCCCTGTTTGCATTCGACGCTGCTTACGAAAGTCAAAATTATCGTACCTTTGA 27345  
QY 285 uLeuAlaSerHisProSerArgGluLysPro----- 295  
Db 27346 CTGACGGTCCCTCGCTGACGAAAGTCCGCGGCTCCGGGGTTGAACCTCACTCCGGG 27405  
QY 296 -----PheGluAspPheLeuProSerHisTyrAsnTyrLeuGlnPh 309  
Db 27406 GCTGTGACGTCGGCTTACCTTCGCAAAATTGTACTGAGGACTACCACGCCACGAGAT 27465  
QY 309 eAlaTyrTyrAsnIleGlyAsnTyrThrGlnAlaGlyCysAlaLysThrTyrLeuLe 329  
Db 27466 TAGGTTCTAC----- 27475  
QY 329 uPhePheProAsnAspGluValMetAsnGlnAsnLeuAlaTyrTyrAlaAlaMetLeuG1 349  
Db 27476 -----GAAGACCAATCCCGCCGCCAAATGCGGAGCTTACCGCTCGGTATTAC 27525  
QY 349 yGluGluHisThrArgSerIleGlyProArgGluSerAlaLysGluTyrArgGlnArgSe 369  
Db 27526 CCAGGGCCACATT-----CTTGCCCAATTGCAAGCCATCAACAAAGCCCGCCAGAGTT 27579  
QY 369 rLeuLeuGluLysGluLeuLeuPheAlaTyrAspValPheGlyIleProPheValAs 389  
Db 27580 TCTGCTACGAAGGGACGG-----GGGTT---TACTTGA 27612  
QY 389 pProAspSerTrpThrProGluGluValIleProLysArgLeuGlnGluLysGlnLysSe 409  
Db 27613 CCCCAGTCCCGCGAGGAGCTCAACCCCAATCCCGCCCGCGAGCCCTATCAGCAGCA 27672  
QY 409 rGluArgGluThrAla 414

Db 27673 GCGCGGGGCCCTTGCT 27688  
RESULT 14  
US-08-735-609-4  
; Sequence 4, Application US/08735609  
; Patent No. 5994132  
; GENERAL INFORMATION:  
; APPLICANT: Chamberlain, Jeffrey S.  
; Amalfitano, Andrea  
; Hauser, Michael A.  
; Kumar-Singh, Rajendra  
; Hartigan-O'Connor, Dennis J.  
; TITLE OF INVENTION: IMPROVED ADENOVIRUS VECTORS  
; NUMBER OF SEQUENCES: 15  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Medlen & Carroll, LLP  
; STREET: 220 Montgomery Street, Suite 2200  
; CITY: San Francisco  
; STATE: California  
; COUNTRY: United States Of America  
; ZIP: 94104  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patentin Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/735,609  
; FILING DATE: 23-Oct-1996  
; CLASSIFICATION: <Unknown>  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Ingolia, Diane E.  
; REGISTRATION NUMBER: 40,027  
; REFERENCE/DOCKET NUMBER: UM-02484  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (415) 705-8410  
; TELEFAX: (415) 397-8338  
; INFORMATION FOR SEQ ID NO: 4:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 34303 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: double  
; TOPOLOGY: linear  
; MOLSCULE TYPE: other nucleic acid  
; DESCRIPTION: /desc = "DNA"  
; SEQUENCE DESCRIPTION: SEQ ID NO: 4:  
US-08-735-609-4  
Alignment Scores:  
Pred. No.: 2,34 Length: 34303  
Score: 115.50 Matches: 96  
Percent Similarity: 30.94% Conservative: 42  
Best Local Similarity: 21.52% Mismatches: 146  
Query Match: 2.98% Indels: 162  
DB: 2 Gaps: 20  
US-10-045-815-4 (1-736) x US-08-735-609-4 (1-34303)  
QY 42 GlyThrAlaAlaTyrAlaArgGlyAspTrpProGlyValValLeuSerMetGluArgAla 61  
Db 26620 GGAGATGACGCGCAAGCTAGAGAAAC-----ATTGCACTACACCTTTTCGACAGG 26670  
QY 62 LeuArgSer-----ArgAlaAlaLeuArgAlaLeuArgLeuArg 74  
Db 26671 CTACGTACCCAGGCTGCAAGATCCNACGTGGAGCTCTGCAACCTGGTCTCCTA--- 26727  
QY 75 CysArgThrGlnCysAlaAlaAspPheProTrpGluLeu-----AspProAspTrpSer 92  
Db 26728 -----CCTTGGAAATTTTGCAGAAACCGCTTGGGCA 26760  
QY 93 ProSerPro-----AlaGlnAlaSerGlyAlaGlyAlaLeuArgAsp----- 106  
Db 26761 AAACGTGCTTCATTCCACGCTCAAGGGGAGGCGCGCGGACTACGTCGGGACTGCGT 26820



Qy 93 ProSerPro-----AlaGlnAlaSerGlyAlaGlyAlaLeuArgAsp----- 106  
Db 26761 AAACGTGCTTCAATCCACGCTCAAGGCGGAGCGCGCGGACTACGTCGCGACGTGCGT 26820  
Qy 107 -----LeuSerPhePheGlyGlyLeuLeuArgAlaAla-CysLeuAr 121  
Db 26821 TTACTTATTCTATGCTACACCTGGCAGCGCCATGGGGTTGGCAGCAGTCTTGA 26880  
Qy 121 gArgCysLeuGlyProProAlaAlaHisSerLeuSerGluGluMetGluLeu----- 138  
Db 26881 GGAGTGC-----AACTCAAGAGAGCTGCAGAACTGTCTAAAGCA 26919  
Qy 139 -----GluPheArgLysArgSerProTyrAsnTyrLeuGlnVa 151  
Db 26920 AAACCTGAAGACCTATGGAGCGCTTCACAGCGCTCC----- 26959  
Qy 151 lAlaTyrPheLysIleAsnLysLeuGluLysAlaValAlaAlaAlaHisThrPhePheVa 171  
Db 26960 -----GTGGCGCGCACCTGGCGGACAT 26982  
Qy 171 lGlyAsnProGluHisMet-----GluMetGlnGlnAsnLeuAspTyrTyrGlnThrMe 189  
Db 26983 CATTTTCCCGAAGCGCTGTAAACCCCTGCAACAGGGTCTG----- 27025  
Qy 189 tSerGlyValLysGluAlaAspPheLysAspLeuGluThrGlnProHisMetGlnGluPh 209  
Db 27026 -----CCAGACTTCACCACTCAAGCATGTTGCAGAACTT 27060  
Qy 209 eArgLeuGlyValArgLeuTyrSerGluGluProGlnGluAlaValProHisLeuGl 229  
Db 27061 TAGGAACCTTATCCTA-----GAGCGCTCAGGAATCTTGGCCGCCACCTG 27105  
Qy 229 uAlaAlaLeuGlnGluTyrPheVal-----AlaTyrGluGluCysArgAlaLeuCysGl 247  
Db 27106 CTGTGACTTCCTAGCAGCTTTGTGCCCATTAAGTACGGGAAATGCCCTCGCGCGCTTTG 27165  
Qy 247 uGlyPro-----TyrAspTyrAspGlyTyrAsnTyrLeuGluTyrAsnAlaLeuPheGl 266  
Db 27166 GGGCCACGTCTACCTTCGACGTAGCCAACTACCTTGCCTACCACTCTGCATAATGA 27225  
Qy 266 nAlaIleThrAspHisTyrIleGlnValLeuAsnCysLysGlnAsnCysValThr----- 284  
Db 27226 AGACGTGACGGGTACGGTCTACTGGAGTGTCACTGTGCTGCAACCTATGCACCCCGCA 27285  
Qy 285 -----G1 285  
Db 27286 CCGCTCCCTGTTTGCATTCGACGTCTTAACGAAAGTCAAATTAATCGGTACCTTTGA 27345  
Qy 285 uLeuAlaSerHisProSerArgGluLysPro----- 295  
Db 27346 GCTGACGGGTCCCTGCGCTGACGAAAGTCCGCGGCTCCGGGTTGAAACTCACTCCGGG 27405  
Qy 296 -----PheGluAspPheLeuProSerHisTyrAsnTyrLeuGlnPh 309  
Db 27406 GCTGTGGACGTGGCTTACCTTCGCAANTTTGTACCTGAGGACTACCACGCCACGAGAT 27465  
Qy 309 eAlaTyrTyrAsnIleGlyAsnTyrThrGlnAlaGlyGluCysAlaLysThrTyrLeuLe 329  
Db 27466 TAGGTTCTAC----- 27475  
Qy 329 uPhePheProAsnAspGluValMetAsnGlnAsnLeuAlaTyrTyrAlaAlaMetLeuGl 349  
Db 27476 -----GAAGACCAATCCCGCCCGCCAAATGCGGAGTTACCGCTCGCTCATAC 27525  
Qy 349 yGluGluHisThrArgSerIleGlyProArgGluSerAlaLysGluTyrArgGlnArgSe 369  
Db 27526 CCAGGGCCACATT-----CTTGGCAATTTGAGCCCATCAACAAAGCCGCCAAGAGTT 27579  
Qy 369 rLeuLeuGluLysGluLeuLeuPheAlaTyrAspValPheGlyIleProPheValas 389  
Db 27580 TCTGCTACGAAAGGACCG-----GGGGTT---TACTTGA 27612

Search completed: July 18, 2004, 18:12:21  
Job time : 947 secs

Qy 389 pProAspSerT-pThrProGluValIleProLysArgLeuGlnGluLysGlnLysSe 409  
Db 27613 CCCCAGTCCGCGGAGGAGCTCAACCAATCCCCCGCGCGGAGCCCTATCAGCAGCA 27672  
Qy 409 rGluArgGluThrAla 414  
Db 27673 GCGCGGGGCCCTTGCT 27688

This Page Blank (uspto)

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - nucleic search, using frame\_plus\_p2n model

Run on: July 18, 2004, 10:20:10 ; Search time 977 Seconds  
(without alignments)  
3200.280 Million cell updates/sec

Title: US-10-045-815-4

Perfect score: 736

Sequence: 1 MAVRALKLLTLLAVAAAS.....PPEPAQESLSGSSEKPKDEL 736

Scoring table:

OLIGO  
Xgapop 60.0 , Xgapext 60.0  
Xgapop 60.0 , Xgapext 60.0  
Fgapop 6.0 , Fgapext 7.0  
Delop 6.0 , Delext 7.0

Searched: 3373663 seqs, 2124099041 residues

Word size: 1

Total number of hits satisfying chosen parameters: 6745146

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Command line parameters:

-MODEL=frame+ p2n.model -DEV=xlp  
-Q=/cgn2\_1/USPTO.spool\_p/US10045815/runat\_14072004\_123118\_17231/app\_query.fasta\_1.903  
-DB=N Geneseq\_29Jan04 -OMT=fastap -SUFFIX=oli.rng -MINMATCH=0.1 -LOOPCL=0  
-LOPEXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=oligo -TRANS=human40.cdi  
-LIST=45 -DOALIGN=200 -THR\_SCORE=quality -THR\_MIN=1 -ALIGN=15 -MODE=LOCAL  
-OUTFMT=ptc -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000  
-USER=US10045815.@CGN\_1\_1\_885@runat\_14072004\_123118\_17231 -NCPU=6 -ICPU=3  
-NO MWAP -LARGQUERY -NEG SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG  
-DEV TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=60 -XGAPEXT=60 -FGAPOP=6  
-FGAEXT=7 -YGAPOP=60 -YGAPEXT=60 -DELOP=6 -DELEXT=7

Database : N\_Geneseq\_29Jan04:.\*  
1: Geneseqn1980s:.\*  
2: Geneseqn1990s:.\*  
3: Geneseqn2000s:.\*  
4: Geneseqn2001as:.\*  
5: Geneseqn2001bs:.\*  
6: Geneseqn2002s:.\*  
7: Geneseqn2003as:.\*  
8: Geneseqn2003bs:.\*  
9: Geneseqn2003cs:.\*  
10: Geneseqn2004s:.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	736	100.0	2600	3	AAC64725 Human tum
2	634	86.1	2242	6	Abz11346 Human pol
3	433	58.8	2211	6	Aas17573 DNA encod
4	433	58.8	2563	5	Aaf93800 Human CDN
5	433	58.8	2585	4	Aah14589 Human CDN
6	407	55.3	2524	4	Aah14513 Human CDN
7	407	55.3	2577	4	Aad08489 Human sec
8	390	53.0	2615	4	Aad08507 Human sec

9	380	51.6	2753	6	ABK94910	Abk94910 Human nov
10	380	51.6	2753	6	ABK94933	Abk94933 Human nov
11	360	48.9	2829	3	AAC64724	AAC64724 Human tum
12	295	40.1	2993	4	AAH16288	AAH16288 Human CDN
13	275	37.4	2127	6	AAS17572	Aas17572 DNA encod
14	244	33.2	2152	6	ABZ11345	Abz11345 Human pol
15	219	29.8	1309	6	ABZ11344	Abz11344 Human pol
16	188	25.5	1607	6	ABQ76716	Abq76716 Human leu
17	179	24.3	550	7	ACA57060	ACA57060 Human adi
18	172	23.4	1595	3	AAC76943	AAC76943 Human ORF
19	103	14.0	682	4	AAI96734	AAI96734 Human neu
20	101	13.7	652	4	AAI96735	AAI96735 Human neu
21	101	13.7	738	4	AAH06612	AAH06612 Human CDN
22	101	13.7	774	4	AAH06885	AAH06885 Human CDN
23	91	12.4	2322	3	AAC64727	AAC64727 Mouse tum
24	91	12.4	2416	3	AAC64726	AAC64726 Mouse tum
25	87	11.8	465	8	ACH42115	ACH42115 Human foe
26	80	10.9	703	4	AAH06792	AAH06792 Human CDN
27	73	9.9	711	4	AAI96735	AAI96735 Human neu
28	58	7.9	569	4	AAH11909	AAH11909 Human CDN
29	56	7.6	681	5	AAS80583	AAS80583 DNA encod
30	42	5.7	554	4	AAH11825	AAH11825 Human sec
31	37	5.0	162	3	AAC28769	AAC28769 Human sec
32	28	3.8	535	5	AAF94143	AAF94143 Primer sp
33	26	3.5	504	8	ACH28054	ACH28054 Human adu
34	18	2.4	462	8	ACH23862	ACH23862 Human adu
35	17	2.3	60	6	ABN40277	ABN40277 Human spl
36	14	1.9	416	3	AAH30421	AAH30421 Human col
37	13	1.8	554	9	ACF79597	ACF79597 Human pro
38	13	1.8	649	9	ACF79593	ACF79593 Human pro
39	13	1.8	684	9	ACF79589	ACF79589 Human pro
40	13	1.8	685	9	ACF79590	ACF79590 Human pro
41	13	1.8	793	9	ACF79586	ACF79586 Human pro
42	13	1.8	849	8	ADB83188	ADB83188 Human CDN
43	13	1.8	899	9	ACF79594	ACF79594 Human pro
44	13	1.8	940	9	ACF79595	ACF79595 Human pro
45	13	1.8	1023	4	AAS27088	Aas27088 CDNA enco

ALIGNMENTS

RESULT 1  
AAC64725

ID AAC64725 standard; cDNA; 2600 BP.

XX AAC64725;

AC AAC64725;

XX 27-FEB-2001 (first entry)

XX Human tumour suppressor Gros1-S encoding cDNA SEQ ID NO:3.

XX Tumour suppressor; Gros1-L; Gros1-S; cell proliferation; regulation;

XX cancer; cytostatic; gene therapy; ss.

XX Homo sapiens.

XX WO200065047-A1.

XX 02-NOV-2000.

XX 26-APR-2000; 2000WO-JP002731.

XX 26-APR-1999; 99JP-00118806.

XX (CHUG-) CHUGAI RES INST MOLECULAR MEDICINE INC.

XX Wadhwa R, Sugihara T, Yoshida A;

XX WPI; 2000-687340/67.

XX P-PSDB; AAB36392.

XX Full-length tumor suppressor genes encoding Gros1-L, Gros1-S and mouse  
PT homologs participating in regulation of cell proliferation, useful in

PT development of preventives and remedies of cancer.

XX Claim 1; Page 70-77; 114pp; Japanese.

XX The present sequence encodes the human tumour suppressor designated Gros1  
 CC -S. Gros1-L and Gros1-S have cytotostatic activity and can be used in gene  
 CC therapy. Gros1-L and Gros1-S genes are useful in the development of drugs  
 CC used to treat and prevent cancer

XX SQ Sequence 2600 BP; 608 A; 757 C; 722 G; 513 T; 0 U; 0 Other;

#### Alignment Scores:

Pred. No.:	0	Length:	2600
Score:	736.00	Matches:	736
Percent Similarity:	100.00%	Conservative:	0
Best Local Similarity:	100.00%	Mismatches:	0
Query Match:	100.00%	Indels:	0
DB:	3	Gaps:	0

US-10-045-815-4 (1-736) x AAC64725 (1-2600)

QY	1	MetAlaValArgAlaLeuLysLeuThrThrLeuLeuAlaValAlaAlaSer	20
DB	52	ATGGCGGTACGGCGTGTGAAGCTGTGACCACTGCTGGCTGCTGGCGCGCTGCCCTCC	111
QY	21	GlnAlaGluValGluSerGluAlaGlyTrpGlyMetValThrProAspLeuPheAla	40
DB	112	CAAGCCGAGGTGAGTCCGAGCGAGATGGGCGATGCTGATCTGCTCTTGGCC	171
QY	41	GluGlyThrAlaAlaValAlaArgGlyAspTrpProGlyValValLeuSerMetGluArg	60
DB	172	GAGGGGACCGGACCTACGCGCGGGGACTGGCGCGGGTGGTCTTGAGCATGGAACGG	231
QY	61	AlaLeuArgSerArgAlaAlaLeuArgAlaLeuArgCysArgThrGlnCysAla	80
DB	232	GGCGTGGCGTCCGGGCGAGCCCTCGCGCCCTTGGCTGGCTGGCGACCCAGTGGCC	291
QY	81	AlaAspPheProTrpGluLeuAspProAspTrpSerProSerProAlaGlnAlaSerGly	100
DB	292	GCCGACTTCCCGTGGGAGCTGGACCCCGACTGGTCCCGACCGCGCCAGGCGCTCGGGC	351
QY	101	AlaGlyAlaLeuArgAspLeuSerPhePheGlyGlyLeuLeuArgAlaAlaCysLeu	120
DB	352	GCCGCGCGCTCGCGGACCTAGACTTCTCGGGGGCCCTTGGCTGGCTGGCTGGCTGGCTG	411
QY	121	ArgArgCysLeuGlyProProAlaAlaHisSerLeuSerGluGluMetGluLeuGluPhe	140
DB	412	CGCGCTGCTCGGGCGCGCGCGCGCCACTCGCTCAGCGAGAGATGGAGTTC	471
QY	141	ArgLysArgSerProTrpAsnTyrLeuGlnValAlaTyrPheLysIleAsnLysLeuGlu	160
DB	472	CGCAAGCGGAGCCCCACCACTACCTACCTGAGTCCGCTACTTCAAGATCAACAGTTGGAG	531
QY	161	LysAlaValAlaAlaAlaHisThrPheValGlyAsnProGluHisMetGluMetGln	180
DB	532	AAAGCTGTGTGAGACACACACTTCTTCTGGGCAATCTCTGAGCACTGGAATGCGAG	591
QY	181	GlnAsnLeuAspTyrTyrGlnThrMetSerGlyValLysGluAlaAspPheLysAspLeu	200
DB	592	CAGAACCTTAGACTATTACCAACCACTGTCTGGATGAGAGGCGGCACTTCAAGATCTT	651
QY	201	GluThrGlnProHisMetGlnGluPheArgLeuGlyValArgLeuTyrSerGluGln	220
DB	652	GAGACTCAACCCCATATGCAAGATTTTGCATGGGAGTGGCACTCTACTCAGAGGAACAG	711
QY	221	ProGlnGluAlaValProHisLeuGluAlaAlaLeuGlnGluTyrPheValAlaTyrGlu	240
DB	712	CCACAGGAAGCTGTGGCCCACTAGAGCGCGCGCTGCAAGAACTATTGTGGCGCTATGAG	771
QY	241	GluCysArgAlaLeuCysGluGlyProTyrAspTyrAspGlyTyrAsnTyrLeuGluTyr	260
DB	772	GAGTGGCGTGGCCCTCGAAGGGCCCTATGACTACGATGGCTACCACTACCTTGGATAC	831

QY	261	AsnAlaAspLeuPheGlnAlaIleThrAspHisTyrIleGlnValLeuAsnCysLysGln	280
DB	832	AACGCTGACCTCTCCAGGCCATCACAGATCATATCATCCAGTCTCTCACTGTAAGCAG	891
QY	281	AsnCysValThrGluLeuAlaSerHisProSerArgGluLysProPheGluAspPheLeu	300
DB	892	AACGTGTACGAGGCTTCTTCCCAAGTCGAGAGAGCCCTTTGAGAGCTTCTCTC	951
QY	301	ProSerHisTyrAsnTyrLeuGlnPheAlaTyrTyrAsnIleGlyAsnTyrThrGlnAla	320
DB	952	CCATCGCATTAATAATCTGCGAGTTGGCTACTATAACATTTGGGAATTATACAAAGCT	1011
QY	321	GlyGluCysAlaLysThrTyrLeuLeuPhePheProAsnAspGluValMetAsnGlnAsn	340
DB	1012	GGTGAATGTGCCAAGACCTATCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTT	1071
QY	341	LeuAlaTyrTyrAlaAlaMetLeuGlyGluGluHisThrArgSerIleGlyProArgGlu	360
DB	1072	TTGGCTATTATGCGCTATGCTTGGAGAGAACACACACAGATCCATCGGCCCGCTGAG	1131
QY	361	SerAlaLysGluTyrArgGlnArgSerLeuLeuGluLysGluLeuLeuPhePheAlaTyr	380
DB	1132	AGTGCACAGGAGTACCGACAGCGAAGCTTCTTGGAAAAAGAACTCTTCTTCTTCTTCTT	1191
QY	381	AspValPheGlyIleProPheValAspProAspSerTrpThrProGluGluValIlePro	400
DB	1192	GATGTTTTTGGAAATCCCTTCTTGGATCOGGATTCTGAGCTCCAGAGAGAGTATCC	1251
QY	401	LysArgLeuGlnGluLysGlnLysSerGluArgGluThrAlaValArgIleSerGlnGlu	420
DB	1252	AAGAGATTCAAGAGAGAAACAGAAAGTCAGAACGGGAAACAGCGTACGATCTCCAGGAG	1311
QY	421	IleGlyAsnLeuMetLysGluIleGluThrLeuValGluLysThrLysLysSerLeu	440
DB	1312	ATTGGGAACCTTATGAGGAATCGAGACCTTGTGGAGAGAGAACCAAGAGTCACTG	1371
QY	441	AspValSerArgLeuThrArgGluGlyProLeuLeuTyrGluGlyIleSerLeuThr	460
DB	1372	GATGTGAGCAGACTGACCGGAGAGTGGCCCTCTGTATGTAAGGCGATCAGTCTCACC	1431
QY	461	MetAsnSerLysLeuLeuAsnGlyTyrGlnArgValValMetAspLysValIleSerAsp	480
DB	1432	ATGAACCTCAAACTCTGAATGGTTACAGCGGGGGTGGATGGAGCGCGTAACTCTCTGAC	1491
QY	481	HisGluCysGlnGluGlnArgLeuThrAsnValAlaAlaThrSerGlyAspGlyTyr	500
DB	1492	CACGAGTGTGAGAGCTGCAGAGACTGACCAATGTGGCAGCAACCTCAGGAGATGGCTAC	1551
QY	501	ArgGlyGlnThrSerProHisThrProAsnGluLysPheTyrGlyValThrValPheLys	520
DB	1552	CGGGGTGAGACCTCCCACTACTCCCAATGAAAGTTCTATGGTGTCTCTCTCTCTCTCT	1611
QY	521	AlaLeuLysLeuGlyGlnGluLysValProLeuGlnSerAlaHisLeuTyrTyrAsn	540
DB	1612	GGCTCAAGCTGGGCGAGAGGCAAGTCTCTCTGAGAGTGGCCCACTCTACTACAC	1671
QY	541	ValThrGluLysValArgArgIleValGluSerTyrPheArgLeuAspThrProLeuTyr	560
DB	1672	GTGCGGAGAAAGTGGCGCATCATCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT	1731
QY	561	PheSerTyrSerHisLeuValCysArgThrAlaIleGluGluValGlnAlaGluArgLys	580
DB	1732	TTTTTCT	1791
QY	581	AspAspSerHisProValHisValAspAsnCysIleLeuAsnAlaGluThrLeuValCys	600
DB	1792	GATGATAGTATCCAGTCCAGTCCAGTCCAGTCCAGTCCAGTCCAGTCCAGTCCAGTCC	1851
QY	601	ValLysGluProProAlaTyrThrPheArgAspTyrSerAlaIleLeuTyrLeuAsnGly	620
DB	1852	GTCAAGAGAGCCCCAGGCTTACCTTCCGAGCTACAGGCGCATCTTCTTCTTCTTCTT	1911
QY	621	AspPheAspGlyGlyAsnPheThrGluLeuAspAlaLysThrValThrAlaGlu	640



1912 GACTTCGATCGCGAACTTTATTTTCACTGAACCTGATCCCAAGACCGTGACGGCAGAG 1971  
641 ValClnProGlnCysGlyArgAlaValGlyPheSerSerGlyThrGluAsnProHisGly 660  
1972 GTGACGCTCAGTGTGAAGAGCGTGGGATTCCTTCAGGCACTGAAACCCACATGGA 2031  
661 VallysAlaValThrArgGlyGlnArgCysAlaIleAlaLeuTrpPheThrLeuAspPro 680  
2032 GTGAAGCTGTTCACACGGGCGAGCGGTGTGCCATCGCCCTGTGTTCACCTGGACCT 2091  
681 ArgHisSerGluArgAspArgValGlnAlaAspIleValLysMetLeuPheSerPro 700  
2092 CGACACACGCGAGCGGACGAGGTGCAGGACGATGACCTGTGAGATGCTCTTCAGGCCA 2151  
701 GluGluMetAspLeuSerGlnGlnProLeuAspAlaGlnGlnGlyProProGluPro 720  
2152 GAAGAGATGGACCTCTCCAGGAGCAGCCCTGGATGCCAGCAGGGCCCCCGAACCT 2211  
721 AlaGlnGluSerLeuSerGlySerGluSerLysProLysAspGluLeu 736  
2212 GCACAGAGTCTCTCAGGCACTGAATCGAAGCCCAAGGATGAGCTA 2259

## RESULT 2

ABZ11346

ID ABZ11346 standard; cDNA; 2242 BP.

XX AC

XX AC

XX XX 20-JAN-2003 (first entry)

XX DT

XX DE

XX XX Human polynucleotide SEQ ID NO 228.

XX KW Human; genome mapping; gene therapy; food supplement; virus; fungus;

XX KW cell-proliferative disorder; neurodegenerative disease; bacterial;

XX KW Parkinson's disease; Alzheimer's disease; autoimmune disease;

XX KW multiple sclerosis; diabetes; genetic disorder; wound; burn; infection;

XX KW arthritis; cytostatic; immunomodulator; neurotropic; dermatological;

XX KW antiparkinsonian; antidiabetic; immunosuppressive; haematological;

XX KW haemostatic; vulnery; fungicide; antibacterial; virucide; protozoacide;

XX KW antiarthritic; gene; ss.

XX OS

XX XX Homo sapiens.

XX XX WO200270539-A2.

XX PN

XX PD

XX XX -12-SEP-2002.

XX XX 05-MAR-2002; 2002WO-US005095.

XX XX 05-MAR-2001; 2001US-00799451.

XX PR

XX XX (HYSE-) HYSEQ INC.

XX PA

XX PI Tang YT, Zhou P, Goodrich RW, Asundi V, Zhang J, Zhao QA, Ren P;

XX PI Xue AJ, Yang Y, Ma Y, Yamazaki V, Chen R, Wang Z, Ghosh M;

XX PI Wehrman T, Wang J, Wang D, Drmanac RT;

XX XX WPI: 2002-759812/82.

XX DR P-PSDB; ABP69129.

XX XX

XX PT New polynucleotides comprising sequences assembled from expressed

XX PT sequence tags (ESTs), useful for treating cell-proliferative,

XX PT neurodegenerative, autoimmune, genetic, myeloid or lymphoid, or platelet

XX PT or coagulation disorders.

XX XX

XX PS Claim 1: SEQ ID NO 228; 1012pp + Sequence Listing; English.

XX XX

XX CC The invention relates to an isolated polynucleotide (1) comprising a

XX CC nucleotide sequence selected from any of 948 sequences (ABZ11119-

XX CC ABZ12066) or their mature protein coding portion, active domain coding

XX CC protein or complementary sequences. The polynucleotides are useful for

XX CC identifying expressed genes or for physical mapping of human genome. The

XX CC

CC encoded polypeptides (ABP68902-ABP69849) are useful as molecular weight  
CC markers, as a food supplement, for generating antibodies, in medical  
CC imaging, screening and diagnostic assays and for treating cell-  
CC proliferative disorders (cancer), neurodegenerative diseases (Parkinson's  
CC or Alzheimer's disease), autoimmune diseases (multiple sclerosis,  
CC diabetes, lupus) genetic disorders, myeloid or lymphoid disorders,  
CC platelet or coagulation disorders, wound, burns, incision, ulcers, liver  
CC or lung fibrosis, infections (bacterial, viral, fungal, parasitic),  
CC arthritis, etc. Note: The sequence data for this patent did not form part  
CC of the printed specification, but was obtained in electronic format  
CC directly from WIPO at ftp.wipo.int/pub/published\_pct\_sequences

XX XX

SQ Sequence 2242 BP; 518 A; 655 C; 625 G; 444 T; 0 U; 0 Other;

## Alignment Scores:

Pred. No.:	0	Length:	2242
Score:	634.00	Matches:	734
Percent Similarity:	99.46%	Conservative:	0
Best Local Similarity:	99.46%	Mismatches:	2
Query Match:	86.14%	Indels:	4
DS:	6	Gaps:	0

US-10-045-815-4 (1-736) x ABZ11346 (1-2242)

QY 1 MetAlaValArgAlaLeuLysLeuLeuThrThrLeuLeuAlaValAlaAlaLaser 20

DB 32 ATGGCGGTACGCGCGTGAAGCTGTGACCACTGTGGCTGTGGCCCTGCCTCC 91

QY 21 GlnAlaGluValGluSerGluAlaGlyTrpGlyMetValThrProAspLeuPheAla 40

DB 92 CAAGCCGAGTGCAGTCCGAGGCGATGGGCATGGTGACCCCTGATCTCTTCGCG 151

QY 41 GluGlyThrAlaAlaTyrAla-ArgGlyAspTrpProGlyValValLeuSerMetGluAr 60

DB 152 GAGGGAGCCGAGCCCTACTC-CGCGGGGACTTGGCCCGGGTGTCTCAGCATGGAACG 210

QY 60 gAlaLeuArgSerArgAlaAlaLeuArgAlaLeuArgLeuArgCysArgThrGlnCysAl 80

DB 211 GCGCTGCGCTCCCGGCGAGCCCTCCGCGCCCTTGGCTGCGGTGCGCACCGAGTGC 270

QY 80 aAlaAspPheProTrpGluLeuAspProAspTrpSerProSerProAlaGlnAlaSerGl 100

DB 271 CGCGAGTTCGCGTGGGAGCTGACCCCGACTGGTCCCCAGCCCGGCCCGGCTCGGG 330

QY 100 YAlaGly-AlaLeuArgAspLeuSerPhePheGlyGlyLeuLeuArgArgAlaAlaCysL 120

DB 331 CGCGCG-CGCGCTCGCGGAGCTTCTTCGGGGGCTTCTGCGTGGCTGCGCTGCGC 389

QY 120 euArgArgCysLeuGlyProProAlaAlaHisSerLeuSerGluGluMetGluLeuGluP 140

DB 390 TGGCGCGCTGCGTGGGCGCGCGCGGCGGCTGCTCAGCAGGAGATGGAGCTGGAGT 449

QY 140 heArgLysArgSerProTyrAsnTyrLeuGlnValAlaTyrPheLysIleAsnLysLeuG 160

DB 450 TCGCAGAGCGGAGCCCTACACTACTCGAGTTCGCTTCTCAAGATCAACAAGTTGG 509

QY 160 lUlysAlaValAlaAlaAlaHisThrPhePheValGlyAsnProGluHisMetGluMetG 180

DB 510 AGAAAGCTGTGTGCGAGCACACACCTTCTTCGTGGGCAATCTCTGAGCATGGAATGC 569

QY 180 lGlnAsnLeuAspTyrTyrGlnThrMetSerGlyValLysGluAlaAspPheLysAspL 200

DB 570 AGCAGAACCTTAGACTATTCAAAACCATGTCTGAGTGAAGAGGCGCGACTTCAAGATC 629

QY 200 euGluThrGlnProHisMetGlnGluPheArgLeuGlyValArgLeuTyrSerGluGluG 220

DB 630 TTGAGACTCAACCCCATATGCAAGAAATTCGACTGGGAGTGGGACTCTACTCAGAGAAC 689

QY 220 lnrGlnGluAlaValProHisLeuGluAlaAlaLeuGlnGluTyrPheValAlaTyrG 240

DB 690 AGCCACAGGAAGTGTGCGCCCACTAGAGCGCGGCTGCAAGAAATCTTTTGGCCCTATG 749

QY 240 luGluCysArgAlaLeuCysGluGlyProTyrAspTyrAspGlyTyrAsnTyrLeuGluT 260

750 AGAGTGGCGTCCCTCTCGAAGGCGCCCTATGACTACGATGCCTACACTACCTTGGAT 809  
 260 YrAsnAlaAspLeuPheGlnAlaIleThrAspHisTyrIleGlnValLeuAsnCysLysG 280  
 810 ACACGCTGACCTTTCAGGCGCATTACAGATCATATCATCCAGTCTCTCACTGTAAGC 869  
 280 InAsnCysValThrGluLeuAlaSerHisProSerArgGluLysProPheGluAspPheL 300  
 870 AGAAGCTGTCTACGAGGCTTCTCCACCCCAAGTCGAGAGAGCCCTTGAAGACTTCC 929  
 300 euProSerHisTyrAsnTyrLeuGlnPheAlaTyrTyrAsnIleGlyAsnTyrThrGlnA 320  
 930 TCCATCGCATTAATATATCTGCAAGTTGCTTACTATAAATTTGGAAATATATACAAAG 989  
 320 laGlyGluCysAlaLysThrTyrLeuLeuPhePheProAsnAspGluValMetAsnGlnA 340  
 990 CTGGTGAATGTGCCAAGACCTATCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 1049  
 340 snLeuAlaTyrTyrAlaAlaMetLeuGlyGluGluHisThrArgSerIleGlyProArgG 360  
 1050 AFTTGGCTTATATGACGATATGCTTGGAGAGAAACACACAGATCCATCGGCCCCCGTG 1109  
 360 luSerAlaLysGluTyrArgGlnArgSerLeuLeuGluLysLeuLeuPhePheAlaT 380  
 1110 AGAGTGCCAGAGTACCGACAGCGAGCCCTACTCGAAGAAAGACTGCTTTCTCTCGCTT 1169  
 380 YrAspValPheGlyIleProPheValAspProAspSerTyrThrProGluGluValIleP 400  
 1170 ATGATGTTTTTGGAAATCTCTTGTGTGGATTCGGATTCATGACTCCAGAGAGTGAATC 1229  
 400 roLysArgLeuGlnGluLysGlnLysSerGluArgGluThrAlaValArgIleSerGlnG 420  
 1230 CCAAGAGATTGCAAGAGAAACAGAAAGTCAAGACGGGAAACAGCCGTACGCAATCTCCAGG 1289  
 420 luIleGlyAsnLeuMetLysGluIleGluThrLeuValGluGluLysThrLysGluSerL 440  
 1290 AGATTGGGAACCTTATGAAGAAATCGAGACCCCTTGTGGAAGAGAGACCAAGAGGTCAC 1349  
 440 euAspValSerArgLeuThrArgGluGlyGlyProLeuLeuTyrGluLysIleSerLeuT 460  
 1350 TGGATGTGAGCAGACTGACCCGGGAAGGTGGCCCTCTGTATGAAGGATCATGTCTCA 1409  
 460 hrMetAsnSerLysLeuLeuAsnGlyTyrGlnArgValValMetAspGlyValIleSerA 480  
 1410 CCATGAACCTCCAACTCCTGAATGTTACACAGCGGTGTGATGAGCGGCTATCTCTG 1469  
 480 spHisGluCysGlnGluLeuGlnArgLeuThrAsnValAlaAlaThrSerGlyAspGlyT 500  
 1470 ACCACGAGTGTGAGGAGCTGAGAGACTGACCAATGTGCGACGCAACCTCAGGAGATGGCT 1529  
 500 YrArgGlyGlnThrSerProHisThrProAsnGluLysPheTyrGlyValThrValPheL 520  
 1530 ACCGGGTGACAGCTCCCACTATCCATGAAAGTTCATGTTGTCTACTGCTCTTCA 1589  
 520 ysAlaLeuLysLeuGlyGlnGluLysValProLeuGlnSerAlaHisLeuTyrTyrA 540  
 1590 AAGCCCTCAAGCTGGGGAAGAGCAAAAGTTCTCTGCGAGTGGCCACCTGTACTACA 1649  
 540 snValThrGluLysValArgGlyMetGluSerTyrPheArgLeuAspThrProLeuT 560  
 1650 ACGTGACGGAGAAAGTGGCGCATCATGAGTCTTACTTCCGCTGGATACGCCCTCT 1709  
 560 YrPheSerTyrSerHisLeuValCysArgThrAlaIleGluGluValGlnAlaGluArgL 580  
 1710 ACTTTCTCTCTCATCTGTTGGCGCACTGCGCATCGAAGAGGTCCAGGACAGAGGA 1769  
 580 YrAspAspSerHisProValHisValAspAsnCysIleLeuAsnAlaGluThrLeuValC 600  
 1770 AGGATGATAGTATCCAGTCCAGTGGACACTGCATCTCTGAATGCCGAGACCTCTGTGT 1829  
 600 YsValLysGluProProAlaTyrThrPheArgAspTyrSerAlaIleLeuTyrLeuAsnG 620

1830 GTGTCAAAGAGCCCCCAGCCTACACCTTCGCGACTACAGCGCCATCTTTACCTAAATG 1889  
 620 lyAspPheAspGlyGlyAsnPheTyrPheThrGluLeuAspAlaLysThrValThrAlaG 640  
 1890 GGGACTTCGATGGCGGAACCTTTTATTTCTGACTGATGCCAAGACCGTGACGGCAG 1949  
 640 luValGlnProGlnCysGlyArgAlaValIlePheSerSerGlyThrGluAsnProHisG 660  
 1950 AGGTGACGCTCAGTGTGGAAGAGCCGTGGGATTTCTTTTACGCACTGAAACCCACATG 2009  
 660 lyValLysAlaValThrArgGlyGlnArgCysAlaIleAlaLeuTyrPheThrLeuAsp 680  
 2010 GAGTGAAGGCTGTACACAGGGGCGAGCTGTGCCATCGCCCTGTGTCTTACCCCTGGAC 2069  
 680 roArgHisSerGluArgAspArgValGlnAlaAspAspLeuValLysMetLeuPheSerP 700  
 2070 CTCGACACAGCGAGCGGACAGGGTGCAGGCAGATGACCTGGTGAAGATGCTCTTCAGCC 2129  
 700 roGluGluMetAspLeuSerGlnGlnProLeuAspAlaGlnGlnGlyProProGluP 720  
 2130 CAGAAGAGATGGACCTCTCCAGGAGCAGCCCTGGATGCCAGAGGGGCCCCCGGAC 2189  
 720 roAlaGlnGluSerLeuSerGlySerGlySerLysProLysAspGluLeu 736  
 2190 CTGCACAGAGTCTCTCTCAGGACGATGATCGAAGCCCAAGGATGAGCTA 2239  
 RESULT 3  
 AAS17573  
 ID AAS17573 standard; cDNA; 2211 BP.  
 XX  
 AC AAS17573;  
 DT 26-FEB-2002 (first entry)  
 DE DNA encoding novel secreted protein #2.  
 XX  
 KW Secreted protein; cytostatic; immunosuppressive; vulnery; vaccine;  
 antiinflammatory; neuroprotective; nephrotropic; cardiovascular; human;  
 cancer; autoimmune disease; wound healing disorder; infection;  
 haematopoietic disorder; inflammatory disorder; infertility;  
 neurological disease; psychiatric disease; cardiovascular disease;  
 respiratory disease; renal; gastrointestinal; ss.  
 XX  
 OS Homo sapiens.  
 Key Location/Qualifiers  
 CDS 1..2211  
 /tag= a  
 /product= "Human secreted protein"  
 XX  
 WO200179454-A1.  
 XX  
 PD 25-OCT-2001.  
 XX  
 PF 11-APR-2001; 2001WO-US011797.  
 XX  
 PR 13-APR-2000; 2000US-0196603P.  
 PR 24-APR-2000; 2000US-0199417P.  
 XX  
 PA (SMIK ) SMITHKLINE BEECHAM CORP.  
 PA (SMIK ) SMITHKLINE BEECHAM PLC.  
 XX  
 PI Agarwal P, Murdoch PR, Rizvi SK, Smith RF, Xiang Z;  
 DR WPI; 2002-061975/08.  
 DR P-PSDB; AAU09861.  
 XX  
 PT New secreted proteins or polypeptides, useful for treating e.g. cancer,  
 autoimmune diseases, wound healing disorder, infections, hematopoietic  
 disorders, inflammatory disorders, infertility, cancer.  
 XX  
 PS Claim 2; Page 35-36; 92pp; English.

CC The invention relates to an isolated novel secreted polypeptide (I) and  
CC polynucleotide (II). (I) and (II) are useful for treating cancer,  
CC autoimmune diseases, wound healing disorder, infections, haematopoietic  
CC disorders, inflammatory disorders, infertility, neurological and  
CC psychiatric diseases, cardiovascular diseases, respiratory diseases,  
CC renal diseases, or gastrointestinal diseases. These may also be used to  
CC treat diseases, abnormalities and disorders caused by abnormal  
CC expression, production, function and/or metabolism of the genes, as  
CC vaccines for inducing immunological response in a mammal, and in  
CC screening methods for detecting the effect of added compounds on the  
CC production of mRNA and polypeptide in cells. The polypeptides can be used  
CC as immunogens to produce antibodies immunospecific for the polypeptides,  
CC and to identify membrane-bound or soluble receptors. The polynucleotides  
CC may be used as diagnostic reagents, in chromosome localisation studies,  
CC and in tissue expression studies. The present sequence represents the  
CC coding sequence of novel human secreted protein #2

XX Sequence 2211 BP; 512 A; 549 C; 612 G; 438 T; 0 U; 0 Other;

Alignment Scores:  
Pred. No.: 0 Length: 2211  
Score: 433.00 Matches: 733  
Percent Similarity: 99.19% Conservative: 0  
Best Local Similarity: 99.19% Mismatches: 3  
Query Match: 58.83% Indels: 6  
DB: 6 Gaps: 0

US-10-045-815-4 (1-736) x AAS17573 (1-2211)

Qy 1 MetAlaValArgAlaLeuLysLeuLeuThrThrLeuLeuAlaValAlaAlaSer 20  
Db 1 ATGGCGGTACGGGGTGAAGCTGTGACCATCTGCTGGCTGTGCTGGCGCTGCC 60  
Qy 21 GlnAlaGluValGluSerGluAlaGlyTrpGlyMetValThrProAspLeuPheAla 40  
Db 61 CAAGCCGAGGTCGAGTCCGAGCAGGATGGGCATGGTGTGACGCTGTCTCTGCC 120  
Qy 41 GluGlyThrAlaAlaValArgGlyAspTrpProGlyValValLeuSerMetGluArg 60  
Db 121 GAGGGAGCCGAGCCCTACGCGCGGGGACATGGCCGGGGTGGTCTGAGCATGGAACG 180  
Qy 61 AlaLeuArgSerArgAlaAlaLeuArgAlaLeuArgLeuArgCysAysThrGlnCysAla 80  
Db 181 GCGTCTCGCTCCCGGAGCCCTCGCGCCCTTCGCTCGCTCGCTCGCTCGCTGCC 240  
Qy 81 AlaAspPheProTrpGluLeuAspProAspTrpSerProSerProAlaGlnAlaSerGly 100  
Db 241 GCGGACTTCCCGTGGAGCTGAGACCCCGACTGGTCCCGCCAGCCCGGCTCGGGC 300  
Qy 101 AlaGly-AlaLeuArgAspLeuSerPheGlyGlyLeuLeuArgArgAlaAlaCysLe 120  
Db 301 GCGGC-CGCCCCCGGACCTGAGCTTCTTGGGGGCTTCTGGCTCGCTGCTGCT 359  
Qy 120 uArgArgCysLeuGlyProProAlaAlaHisSerLeuSerGluGluMetGluLeuGluPh 140  
Db 360 GCGCGCTGCTCGGGCGCGCGCGCCACTCGCTCAGCGAAGAGATGAGCTGGAGTT 419  
Qy 140 eArgLysArgSerProTrpAsnTrpLeuGlnValAlaTrpPheLysIleAsnLysLeuG1 160  
Db 420 CCGCAAGCGGAGCCCTCAACTACTCGAGTGCCTCTCTTCAAGATCAACAGTTGGA 479  
Qy 160 uLysAlaValAlaAlaHisThrPheValGlyAsnProGluHisMetGluMetG1 180  
Db 480 GAAGCTGTGTGCGAGCACACACTTCTTCTGGGCAATCTGAGCACATGGAATGCA 539  
Qy 180 nGlnAsnLeuAspTrpTrpGlnThrMetSerGlyValLysGluAlaAspPheLysAspLe 200  
Db 540 GAGAACCTAGACTATTACCAACCATGTCTGGATGAGAGGCGGACTTCAAGGATCT 599  
Qy 200 uGluThrGlnProHisMetGlnGluPheArgLeuGlyValArgLeuTrpSerGluGluG1 220  
Db 600 TGAGACTCAACCCATATCCAGAAATTCGACTGGAGTGGGACTCTACTCAGAGAACCA 659

Qy 220 nProGlnGluAlaValProHisLeuGluAlaAlaLeuGlnGluTrpPheValAlaTrpG1 240  
Db 660 GCCACAGGAAGTGTGCCCCACCTAGAGCGCGCTGCAAGAATACTTTGTGGCTATGA 719  
Qy 240 uGluCysArgAlaLeuCysGluGlyProTrpAspTrpAspGlyTrpAsnTrpLeuGluTr 260  
Db 720 GGAGTGCCGTGCGCTCTGGCAAGGCGCTATGACTACCATGGCTACACTACTTGA 779  
Qy 260 rAsnAlaAspLeuPheGlnAlaIleThrAspHisTrpIleGlnValLeuAsnCysGlyG1 280  
Db 780 CAACGCTGACCTTCTCCAGGCCATCACAGATCATTTACATCCAGGTCCTCACTGAAGCA 839  
Qy 280 nAsnCysValThrGluLeuAlaSerHisProSerArgGluLysProPheGluAspPheLe 300  
Db 840 GAACGTGTGTCAGGAGCTTGTCTCCACCCCAAGTCCAGAGAAGCCCTTTGAAGACTTCT 899  
Qy 300 uProSerHisTrpAsnTrpLeuGlnPheAlaTrpTrpAsnIleGlyAsnTrpGlnAl 320  
Db 900 CCCATCGCATATATATCTGCGTTCGCTTCTATATACATTTGGGAATATTACACAGGC 959  
Qy 320 aGly-GluCysAlaLysThrTrpLeuLeuPhePheProAsnAspGluValMetAsnGlnA 340  
Db 960 TGT-TGAATGTGCCAGACCTATCTCTCTTCTCCCAATCAGCAGGTGATGAACCAAA 1018  
Qy 340 snLeuAlaTrpTrpAlaAlaMetLeuGlyGluGluHisThrArgSerIleGlyProArgG 360  
Db 1019 ATTTGCGCTATTATGCGATGCTTGTGAGAGAAGACACACAGATCCATCGCCCCCGTG 1078  
Qy 360 luserAlaLysGluTrpArgGlnArgSerLeuLeuGlyLysGluLeuLeuPhePheAlaTr 380  
Db 1079 AGAGTCCCAAGGAGTACCGACAGCGAAGCTTACTTGGAAAAAGAACTGCTTTTCTCGCTT 1138  
Qy 380 rAspValPheGlyIleProPheValAspProAspSerTrpThrProGluGluValIleP 400  
Db 1139 ATGATGTTTTGGAAATCTCTTTGTGGATCCGGATTCATGGACTCCAGAAAGATGATTC 1198  
Qy 400 rLysArgLeuGlnGluLysGlnLysSerGluArgGluThrAlaValArgIleSerGlnG 420  
Db 1199 CCAAGAGATTGCAAGAGAAACAGAGTCAAGACGGAAACAGCGGTACGCATCTCCAGG 1258  
Qy 420 lIleGlyAsnLeuMetLysGluIleGluThrLeuValGluGluLysThrLysGluSerL 440  
Db 1259 AGATTGGGAACCTTATGAAGAAATCGAGACCTTGTGGAAGAGAAGACCAAGGAGTCA 1318  
Qy 440 euAspValSerArgLeuThrArgGluGlyGlyProLeuLeuTrpGluGlyIleSerLeu 460  
Db 1319 TGATGTGACGACACTGACCCGGAGAGTGGCCCCCTGCTGTATGAAGCATCAGTCTCA 1378  
Qy 460 hrMetAsnSerLysLeuLeuAsnGlyTrp-GlnArgValValMetAspGlyValIleSer 479  
Db 1379 CCATGAACCTCAAACTCCTGAATGGTTC-CCAGCGGGTGGTGTGATGGACGCGTAATCTCT 1437  
Qy 480 AspHisGluCysGlnGluLeuGlnArgLeuThrAsnValAlaAlaThrSerGlyAspGly 499  
Db 1438 GACCACAGTGTGAGAGCTGCAGACACTGACCAATGTGGCAGCAACCTCAGAGATGGC 1497  
Qy 500 TyrArgGlyGlnThrSerProHisThrProAsnGluLysPheTrpGlyValThrValPhe 519  
Db 1498 TACCGGGGTGAGACCTCCCCACATCTCCCAATGAAAGTTCTTATGGTGTCTGCTCTTC 1557  
Qy 520 LysAlaLeuLysLeuGlnGluLysValProLeuGlnSerAlaHisLeuTrpTrp 539  
Db 1558 AAAGCCCTCAAGCTGGGGCAAGGCAAGATTTCTCTGCAAGTGGCCCACTGTACTAC 1617  
Qy 540 AsnValThrGluLysValArgArgIleMetGluSerTrpPheArgLeuAspThrProLeu 559  
Db 1618 AACGTGACGAGAGAGTGGCGGCATCATGAGTCTCTACTTCCGCTGGATAGACGCCCTC 1677  
Qy 560 TyrPheSerTrpSerHisLeuValCysArgThrAlaIleGluValGlnAlaGluArg 579  
Db 1678 TACTTTTCTTACTCTCTATCTGGTGGCGCACCTGCCATCGAAGAGGTCCAGGAGAGG 1737  
Qy 580 LysAspAspSerHisProValHisValAspAsnCysIleLeuAsnAlaGluThrLeuVal 599



Db 630 TGAGACTCAACCCCATATGCAAGATTTCGACTGGGAGTCGGAATCTACTCAGAGAAC 589  
Qy 220 nProGlnGluAlaValProHisLeuGluAlaLeuGlnGluTyrPheValAlaTyrG 240  
Db 690 GCCACAGAACTGTGCCCCACCTAGAGCGCGCTGCAAGATACTTTGTGGCCATGA 749  
Qy 240 uGluCysArgAlaLeuCysGluGlyProTyrAspTyrAspGlyTyrAsnTyrLeuGluTyr 260  
Db 750 GGAGTGGCCGTGCCCTTCGCAAGCGCCCTATGACTACGATGGCTCAACACTACCTTGA 809  
Qy 260 rAsnAlaAspLeuPheGlnAlaLeuThrAspHisTyrIleGlnValLeuAsnCysLysG 280  
Db 810 CAAGCCTGACCTTTCAGGCCATCACAGATCATTACATCCAGTCTCAACTGTAAACA 869  
Qy 280 nAsnCysValThrGluLeuAlaSerHisProSerArgGluLysProPheGluAspPheLe 300  
Db 870 GAACGTGTCCAGGAGCTTGTCTCCACCCCAAGTCGAGAGAACCCCTTTGAAGACTTCCT 929  
Qy 300 uProSerHisTyrAsnTyrLeuGlnPheAlaTyrTyrAsnIleGlyAsnTyrThrGlnAl 320  
Db 930 CCCATCGCATATATAATATCTGCAAGTTTGGCTACTATAACATTGGGAATATACAGGC 989  
Qy 320 aGly-GluCysAlaLysThrTyrLeuLeuPhePheProAsnAspGluValValMetAsnGlnA 340  
Db 990 TGT-TGAATGTGCCAAGACCTATCTTCTTCTTCCCCAATGACGAGGTGATGAACCAA 1048  
Qy 340 snLeuAlaTyrTyrAlaAlaMetLeuGlyGluGluHisThrArgSerIleGlyProArgG 360  
Db 1049 ATTTGGCCTATTATGCACTATGCTTGGAGAAGAACACACAGATCCATCGGCCCGCTG 1108  
Qy 360 luSerAlaLysGluTyrArgGlnArgSerLeuLeuGluLysGluLeuLeuPhePheAla 380  
Db 1109 AGATGTCAGAGAGTACGACAGCGAAGCCTACTGGAAAGAACTCTCTTCTTCGCTT 1168  
Qy 380 yrAspValPheGlyIleProPheValAspProAspSerTrpThrProGluGluValIleP 400  
Db 1169 ATGATGTTTTTGGAAATCCCTTTCTGGATCCGGATTCATGCACTCCAGAAAGATGATTC 1228  
Qy 400 roLysArgLeuGlnGluLysGlnLysSerGluArgGluThrAlaValArgIleSerGlnG 420  
Db 1229 CCAAGAGATTCCAGAGAAACAGAAAGTCAGAACCGGAAACAGCCGTACGATCTCCCAAG 1288  
Qy 420 luIleGlyAsnLeuMetLysGluIleGluThrLeuValGluGluLysThrLysGluSerL 440  
Db 1289 AGATTGGGAACCTTATGAAGGAATTCGAGACCCCTTGTGGAAGAGAACCAAGAGTCA 1348  
Qy 440 euAspValSerArgLeuThrArgGluGlyGlyProLeuLeuTyrGluGlyIleSerLeuT 460  
Db 1349 TGGATGTGAGCAGACTGACCCGGGAAGGTGCCCCCTGCTGTATGAAGGCATCAGTCTCA 1408  
Qy 460 hrMetAsnSerLysLeuLeuAsnGlyTyr-GlnArgValValMetAspGlyValIleSer 479  
Db 1409 CCATGAATCCAACTCTCTGATGGTTC-CCAGCGGGTGGTGTGATGGAGCGGTAATCTCT 1467  
Qy 480 AspHisGluCysGlnGluLeuGlnArgLeuThrAsnValAlaAlaThrSerGlyAspGly 499  
Db 1468 GACCACGAGTGTCCAGAGCTGCAGAGACTGACCAATGTGGCAGCAACCTCAGAGAGATGC 1527  
Qy 500 TyrArgGlyGlnThrSerProHisThrProAsnGluLysPheTyrGlyValThrValPhe 519  
Db 1528 TACCGGGGTGACAGCTCCCACTACTCCCAATGAAAGTTCTATGTGTGTCATCTGTCTC 1587  
Qy 520 LysAlaLeuLysLeuGlyGlnGluLysValProLeuGlnSerAlaHisLeuTyrTyr 539  
Db 1588 AAAGCCCTCAAGCTGGGGCAAGAGGCAAGTTCCTCTGCAGAGTGCACCACCTGTACTAC 1647  
Qy 540 AsnValThrGluLysValArgIleMetGluSerTyrPheArgLeuAspThrProLeu 559  
Db 1648 AACGTGACGGAGAGGTGGCGCATCATGAGTCTTACTTCCGCTCGATACGCCCTC 1707  
Qy 560 TyrPheSerTyrSerHisLeuValCysArgThrAlaIleGluGluValGlnAlaGluArg 579  
Db 1708 TACTTTTCTACTCTCACTGGTGTGCCACTGCGCATCGAAGAGGTCCAGCGAGAGAGG 1767

Qy 580 LysAspAspSerHisProValHisValAspAsnCysIleLeuAsnAlaGluThrLeuVal 599  
Db 1768 AAGGATGATAGTATCCAGTCCAGTCCAGTCCAGTCCAGTCCAGTCCAGTCCAGTCCAGT 1827  
Qy 600 CysValLysGluProProAlaTyrThrPheArgAspTyrSerAlaIleLeuTyrLeuAsn 619  
Db 1828 TGTGTCAAAGAGCCCGGAGCTTACCTTCCGAGCTACAGGCCATCTTTTACCTAAAT 1887  
Qy 620 GlyAspPheAspGlyGlyAsnPheTyrPheThrGluLeuAspAlaLysThrValThrAla 639  
Db 1888 GGGGACTTCGATGGCGGAAACTTTTATTTTCACTGAACCTGGATGCCAAGACCGTGACGCA 1947  
Qy 640 GluValGlnProGlnCysGlyArgAlaValGlyPheSerSerGlyThrGluAsnProHis 659  
Db 1948 GAGGTGACGCTCAGTGTGAAGAGCCGTGGATTTCTTTCAGGCATCGAAACCCACAT 2007  
Qy 660 GlyValLysAlaValThrArgGlyGlnArgCysAlaIleAlaLeuTrpPheThrLeuAsp 679  
Db 2008 GGAGTGAAGGCTGTCAACAGGGGCGAGCGCTGTGCATCGCCCTGTGGTTTACCCTGGAC 2067  
Qy 680 ProArgHisSerGluArgAspArgValGlnAlaAspAspLeuValLysMetLeuPheSer 699  
Db 2068 CCTCGACACAGCGAGCGGACAGGGTGCAGGCAGATGACCTGTTGAAGATGCTCTTCAGC 2127  
Qy 700 ProGluGluMetAspLeuSerGlnGluProLeuAspAlaGlnGlnGlyProProGlu 719  
Db 2128 CCAGAAAGATGGACCTCTCCAGAGCAGCCCTGGATGCCAGACGGCCCCCGGAA 2187  
Qy 720 ProAlaGlnGluSerLeuSerGlySerGluSerLysProLysAspGluLeu 736  
Db 2188 CTTGCACAAAGAGTCTCTCTCAGGCAGTGAATCGAAGCCCAAGGATGAGCTA 2238

## RESULT 5

AAH14589 standard; cDNA; 2585 BP.

AC AAH14589;

26-JUN-2001 (first entry)

Human cDNA sequence SEQ ID NO:12193.

Human; primer; detection; diagnosis; antisense therapy; gene therapy; ss.

Homo sapiens.

EP1074617-A2.

07-FEB-2001.

28-JUL-2000; 2000BP-00116126.

29-JUL-1999; 99JP-00248036.

27-AUG-1999; 99JP-00300253.

11-JAN-2000; 2000JP-00118776.

02-MAY-2000; 2000JP-00183767.

09-JUN-2000; 2000JP-00241899.

(HELI-) HELIX RES INST.

Ota T, Isogai T, Nishikawa T, Hayashi K, Saito K, Yamamoto J;

Ishii S, Sugiyama T, Wakamatsu A, Nagai K, Otsuki T;

WPI; 2001-318749/34.

Primer sets for synthesizing polynucleotides, particularly the 5602 full-

length cDNAs defined in the specification, and for the detection and/or

diagnosis of the abnormality of the proteins encoded by the full-length

cDNAs.

Claim 8; SEQ ID NO 12193; 2537pp + Sequence Listing; English.

The present invention describes primer sets for synthesising 5602 full-length cDNAs defined in the specification. Where a primer set comprises: (a) an oligo-dr primer and an oligonucleotide complementary to the complementary strand of a polynucleotide which comprises one of the 5602 nucleotide sequences defined in the specification, where the oligonucleotide comprises at least 15 nucleotides; or (b) a combination of an oligonucleotide comprising a sequence complementary to a complementary strand of a polynucleotide which comprises a 5'-end sequence and an oligonucleotide comprising a sequence complementary to a polynucleotide which comprises a 3'-end sequence, where the oligonucleotide comprises at least 15 nucleotides and the combination of the 5'-end sequence/3'-end sequence is selected from those defined in the specification. The primer sets can be used in antisense therapy and in gene therapy. The primers are useful for synthesising polynucleotides, particularly full-length cDNAs. The primers are also useful for the detection and/or diagnosis of the abnormality of the proteins encoded by the full-length cDNAs. The primers allow obtaining of the full-length cDNAs easily without any specialised methods. AAH03166 to AAH13628 and AAH13633 to AAH18742 represent human cDNA sequences; AA895446 to AA895993 represent human amino acid sequences; and AAH13629 to AAH13632 represent polynucleotides, all of which are used in the exemplification of the present invention.

Sequence 2585 BP; 590 A; 759 C; 723 G; 513 T; 0 U; 0 Other;

Alignment Scores:

Pred. No.:	0	Length:	2585
Score:	433.00	Matches:	733
Percent Similarity:	99.19%	Conservative:	0
Best Local Similarity:	99.19%	Mismatches:	3
Query Match:	58.83%	Indels:	6
DB:	48	Gaps:	0

US-10-045-815-4 (1-736) X AAH14589 (1-2585)

Qy	1	MetAlaValArgAlaLeuLysLeuLeuThrThrLeuLeuAlaValValAlaAlaSer	20
Db	53	ATGGGGGTACGGCGCTTGAAGCTGTGACCACACTGTGTGCTCTGTGGCGCTGCTCC	112
Qy	21	GlnAlaGluValGluSerGluAlaClyTrpGlyMetValThrProAspLeuLeuPheAla	40
Db	113	CAAGCGGAGGTCCAGTCCGAGCGAGATGGGCAATGGTGACGCTGATCTGCTTTCGCC	172
Qy	41	GluGlyThrAlaAlaTyrAlaArgGlyAspTrpProGlyValValLeuSerMetGluArg	60
Db	173	GAGGGGACCGCAGCCCTACGCGCGGGGACTTGGCCCGGGGTGCTCTGAGCATGGAAACGG	232
Qy	61	AlaLeuArgSerArgAlaAlaLeuArgAlaLeuArgGluLeuArgCysArgThrGlnCysAla	80
Db	233	CGCGTGTGCGCTCCCGGGCAGCCCTCCGCGCCCTTTCGCGTGGCGCACCCAGATGTGCC	292
Qy	81	AlaAspPheProTrpGluLeuAspProAspTrpSerProSerProAlaGlnAlaSerGly	100
Db	293	GCCGACTTCCCGTGGGAGCTGGACCCGACTGGTGCCCCAGCCGCGCCAGGCTTCGGCG	352
Qy	101	AlaGly-AlaLeuArgAspLeuSerPheGlyGlyLeuLeuArgArgAlaAlaCysLe	120
Db	353	CGCGC- CGCCCTTGGCGCAGCTTCTTTCGGGGGCTTCTCGTTCGGCTGCTGCTGCT	411
Qy	120	uArgArgCysLeuGlyProProAlaAlaHisSerLeuSerGluLeuLeuGluLeuGluPh	140
Db	412	GCGCGCTTGCCTTGGGCGCGCGCGCCCACTCGCTCAGCGAAGATGGAGCTGGAGTT	471
Qy	140	eArgLysArgSerProTyrAsnTyrLeuGlnValAlaTyrPheLysIleAsnLysLeuGln	160
Db	472	CGCAAGCGGAGCGCCCTACAACTACCTGCAGGTTCGCTCTTCAAGATCAACAAGTTGGA	531
Qy	160	uLysAlaValAlaAlaAlaHisThrPhePheValGlyAsnProGluHisMetCluMetCln	180
Db	532	GAAGCTCTTCTCGAGCACACCTTCTTTCGTGGGCAATCTCTGAGCATCGAATAATGCA	591
Qy	180	nGlnAsnLeuAspTyrTyrGlnThrMetSerGlyValLysGluAlaAspPheLysAspLe	200



```
QY 560 TyrPheSerTyrSerHisLeuValCysArgThrAlaLeuGluValGlnAlaGluArg 579
DB 1730 TACTTTCTCTACTCTCATCTGGTGGCGCACTGCCATCGAAGAGGTCCAGCGCAGAGG 1789
QY 580 LysAspSerHisProValHisValAspSerCysIleLeuAenAlaGluThrLeuVal 599
DB 1790 AAGGATGATGATCATCCAGTCCAGTGGACAACTGCATCTCTGAATGCCGAGACCCCTCGTG 1849
QY 600 CysValIysGluProProAlaTyrThrPheArgAspTyrSerAlaIleLeuTyrLeuAsn 619
DB 1850 TGTGTCAAAGAGCCCCCAGCGCTACACCTTCGCGCACTAGAGCGCCATCCCTTACCTAAAT 1909
QY 620 GlyAspPheAspGlyGlyAsnPheTyrPheThrGluLeuAspAlaLysThrValThrAla 639
DB 1910 GGGGACTTCGATGGCGGAACTTTTATTTCACTGAACCTGGATGCCAAGCCGTCGACGGCA 1969
QY 640 GluValGlnProGlnCysGlyArgAlaValGlyPheSerSerGlyThrGluAsnProHis 659
DB 1970 GAGGTGAGCTCTAGTGTGGAGAGCGGTGGATTCCTTCAGCATCTGAARACCCACAT 2029
QY 660 GlyValIysAlaValThrArgGlyGlnArgCysAlaIleAlaLeuTyrPheThrLeuAsp 679
DB 2030 GGAGTGAAGGCTGTCAACAGGGGCGAGCGCTGTGCCATCGCCCTGTGTTCAACCTGGAC 2089
QY 680 ProArgHisSerGluArgAspArgValGlnAlaAspLeuValIysMetLeuPheSer 699
DB 2090 CCTCGACACAGCGAGCGGAGCAGGTGCAGCAGATGACCTGTGTGAAGATGCTCTTCAGC 2149
QY 700 ProGluGluMetAspLeuSerGlnGlnProLeuAspAlaGlnGlnGlyProProGlu 719
DB 2150 CCAGAGAGATGGACCTCTCCAGAGCAGCGCCCTGGATGCCAGCAGGGCCCCCGAA 2209
QY 720 ProAlaGlnGluSerLeuSerGlySerGluSerIysProIysAspGluLeu 736
DB 2210 CCTGCACAAGAGTCTCTCTCAGGAGTGAATCGAAGCCCAAGGATGAGCTA 2260
RESULT 6
AAH14513
ID AAH14513 standard; cDNA; 2524 BP.
XX
AC AAH14513;
XX
DT 26-JUN-2001 (first entry)
XX
DE Human cDNA sequence SEQ ID NO:12044.
XX
KW Human; primer; detection; diagnosis; antisense therapy; gene therapy; ss.
XX
OS Homo sapiens.
XX
PN EPI074617-A2.
XX
PD 07-FEB-2001.
XX
PF 28-JUL-2000; 2000EP-00116126.
XX
PR 29-JUL-1999; 99JP-00248036.
PR 27-AUG-1999; 99JP-00300253.
PR 11-JAN-2000; 2000JP-00118776.
PR 02-MAY-2000; 2000JP-00183767.
PR 09-JUN-2000; 2000JP-00241899.
XX
PA (HELI-) HELIX RES INST.
XX
PI Ota T, Tsogai T, Nishikawa T, Hayashi K, Saito K, Yamamoto J;
PI Ishii S, Sugiyama T, Wakamatsu A, Nagai K, Otsuki T;
XX
DR WPI; 2001-318749/34.
XX
PT Primer sets for synthesizing polynucleotides, particularly the 5602 full-length cDNAs defined in the specification, and for the detection and/or diagnosis of the abnormality of the proteins encoded by the full-length
```

```
PT cDNAs.
XX
PS Claim 8; SEQ ID NO 12044; 2537pp + Sequence Listing; English.
XX
CC The present invention describes primer sets for synthesizing 5602 full-length cDNAs defined in the specification. Where a primer set comprises:
CC (a) an oligo-dr primer and an oligonucleotide complementary to the
CC complementary strand of a polynucleotide which comprises one of the 5602
CC nucleotide sequences defined in the specification, where the
CC oligonucleotide comprises at least 15 nucleotides; or (b) a combination
CC of an oligonucleotide comprising a sequence complementary to the
CC complementary strand of a polynucleotide which comprises a 5'-end
CC sequence and an oligonucleotide comprising a sequence complementary to a
CC polynucleotide which comprises a 3'-end sequence, where the
CC oligonucleotide comprises at least 15 nucleotides and the combination of
CC the 5'-end sequence/3'-end sequence is selected from those defined in the
CC specification. The primer sets can be used in antisense therapy and in
CC gene therapy. The primers are useful for synthesizing polynucleotides,
CC particularly full-length cDNAs. The primers are also useful for the
CC detection and/or diagnosis of the abnormality of the proteins encoded by
CC the full-length cDNAs. The primers allow obtaining of the full-length
CC cDNAs easily without any specialised methods. AAH03166 to AAH13628 and
CC AAH13633 to AAH18742 represent human cDNA sequences; AAH92446 to AAH95593
CC represent human amino acid sequences; and AAH13629 to AAH13632 represent
CC oligonucleotides, all of which are used in the exemplification of the
CC present invention
XX
SQ Sequence 2524 BP; 583 A; 747 C; 704 G; 490 T; 0 U; 0 Other;
```

```
Alignment Scores:
Pred. No.: 0 Length: 2524
Score: 407.00 Matches: 707
Percent Similarity: 99.16% Conservative: 0
Best Local Similarity: 99.16% Mismatches: 3
Query Match: 55.30% Indels: 6
DB: 4 Gaps: 0
```

US-10-045-815-4 (1-736) x AAH14513 (1-2524)

```
QY 1 MetAlaValArgAlaLeuLysLeuLeuThrThrLeuLeuAlaValAlaAlaSer 20
DB 16 ATGGCGGTACGGCGGTGAAGCTGCTGACACACACTGCTGGCTGTGGTCCCTGCCTCC 75
QY 21 GlnAlaGluValGluSerGluAlaGlyTyrGlyMetValThrProAspLeuLeuPheAla 40
DB 76 CAAGCGAGTTCAGTCCGAGCAGAGATGGGCGATGGTGGCTGATCTGCTTCCTGCC 135
QY 41 GluGlyThrAlaAlaTyrAlaArgGlyAspTyrProGlyValValLeuSerMetGluArg 60
DB 136 GAGGGGACCGCAGCCCTACCGCGCGGGGACTGGCCCGGCTGCTGCTGAGCATGGAACGG 195
QY 61 AlaLeuArgSerArgAlaAlaLeuArgAlaLeuArgLeuArgCysArgThrGlnCysAla 80
DB 196 GCGCTGCGCTCCCGGGCAGCCCTCCGCGCCCTTCGCTGCTGCTGCGCACCAGTGTGCC 255
QY 81 AlaAspPheProTyrGluLeuAspProAspTyrProSerProSerProAlaGlnAlaSerGly 100
DB 256 GCCGACTTCCCGTGGAGCTGGACCCGACTGGTCCCGCAGCCCGCCAGCCCTCGGCG 315
QY 101 AlaGlyAlaLeuArgAspLeuSerPhePheGlyGlyLeuLeuArgArgAlaAlaCysLeu 120
DB 316 GCGCGC-CGCGCTGCGCGACCTGAGCTTCTTCGGGGGCTTCTGCGCTGCGCTGCGCTGCT 374
QY 120 uArgArgCysLeuGlyProProAlaAlaHisSerLeuSerGluGluMetGluLeuGluPh 140
DB 375 GCGCGCTGCTTCGGCGCGCGCCGCGCCACTCGCTCAGCGAGAGAGATGGAGCTGAGTT 434
QY 140 eArgIysArgSerProTyrAsnTyrLeuGlnValAlaTyrPheIlysIleAsnLysLeuGlu 160
DB 435 CGCAAGCGGAGCCCTCACTACTGTCAGTCCGCTTCTCAAGATCAACAGATTGGA 494
QY 160 uLysAlaValAlaAlaAlaHisThrPhePheValGlyAsnProGluHisMetGluMetGlu 180
```



Ds	495	GAAGCTCTGTCGACGACACACCTTCTCGTGGCAATCCTCGAGCACATGGAAATGCA	554	QY	540	AsnValThrGluValArgArgileMetGluSerTyrPheArgLeuAspThrProLeu	559
QY	180	nGlnAsnLeuAspTyrTyrGlnThrMetSerGlyValLysGluAlaAspPheLysAspLe	200	Ds	1633	AAAGTCACGAGAGAGTGGCGGCATCATGGAGTCTACTTCCGCTGGATACGCCCTC	1692
Ds	555	CGAGAACCTTAGACTATACCAACCATCTGTGGAGTGAAGAGCCGACATTCAGAGATCT	614	QY	560	TyrPheSerTyrSerHisLeuValCysArgThrAlaIleGluGluValGlnAlaGluArg	579
QY	200	uGluThrGlnProHisMetGlnGluPheArgLeuGlyValArgLeuTyrSerGluGluGln	220	Ds	1693	TACTTTTCTACTCTCATCTGTGTGCGGCATCGCATCGAAGAGTCCAGCAGAGAGG	1752
Ds	615	TCAGACTCAACCCCATATGCAAGAAATTCGACTGGAGTGGAGTCTACTCTCAGAGGAACA	674	QY	580	LysAspAspSerHisProValHisValAspAsnCysIleLeuAsnAlaGluThrLeuVal	599
QY	220	nProGlnGluAlaValProHisLeuGluAlaIleLeuGlnGluTyrPheValAlaTyrGln	240	Ds	1753	AGGATGATAGTATCATCCAGTCCAGTGGACAACTGCATCTGAATGCCGAGACCCCTGCTG	1812
Ds	675	GCCACAGGAAGCTGTGCCCCACCTAGAGCGCGCTGCAAGAACTCTTTGTGGCTATGA	734	QY	600	CysValLysGluProProAlaTyrThrPheArgAspTyrSerAlaIleLeuTyrLeuAsn	619
QY	240	uGluCysArgAlaLeuCysGluGlyProTyrAspTyrAspGlyTyrAsnTyrLeuGluTyr	260	Ds	1813	TGTGTCAAGAGCCCCCAGCCTACACCTTCCCGCATCTCAGCGCATCTTTTACCTAAT	1872
Ds	735	GGAGTGGCTGCTCTCGAAGGCGCTATGACTACGATGGCTACAACTACCTTTGAGTA	794	QY	620	GlyAspPheAspGlyGlyValAsnPheTyrPheThrGluLeuAspAlaLysThrValThrAla	639
QY	260	rAsnAlaAspLeuPheGlnAlaIleThrAspHisTyrIleGlnValLeuAsnCysLysGln	280	Ds	1873	GGGACTTTCGATGGCGGAACCTTTATTTCACTGAATCGATGCCAAGACCGTGACGCA	1932
Ds	795	CAACGCTGACCTCTTCCAGGCGCATCACAGATCATTACATCCAGGTCTCAACTGTAAAGCA	854	QY	640	GluValGlnProGlnCysGlyArgAlaValGlyPheSerSerGlyThrGluAsnProHis	659
QY	280	nAsnCysValThrGluLeuAlaSerHisProSerArgGluLysProPheGluAspPheLe	300	Ds	1933	GAGTGCAGCCTCAGTGTGGAAAGCCGTGGGATTTCTCTCAGGCATGAAACCCACAT	1992
Ds	855	GNACTGTGTACGGAGCTTGTCTCCACCAAGTCGAGAGAGCCCTTTGAAGACTTCTCT	914	QY	660	GlyValLysAlaValThrArgGlyGlnArgCysAlaIleAlaLeuTyrPheThrLeuAsp	679
QY	300	uProSerHisTyrAsnTyrLeuGlnPheAlaTyrTyrAsnIleGlyAsnTyrThrGlnAl	320	Ds	1993	GGATGGAAGCTGTACACAGGGGGCAGCGCTGTGCATCGCCCTGTGTTTCACTCCCTGGAC	2052
Ds	915	CCCATCGCATATAATATCTGCACTTTGCGCTACTATAAATCGGAATATATACACAGGC	974	QY	680	ProArgHisSerGluArgAspArgValGlnAlaAspAspLeuValLysMetLeuPheSer	699
QY	320	agly-GluCysAlaLysThrTyrLeuLeuPhePheProAsnAspGluValMetAsnGlnA	340	Ds	2053	CCTCGACACAGCGAGCGGACAGGCTGCAGGCAGATGACCTGCTGAAGATGCTCTTCAGC	2112
Ds	975	TGT-TGAATGTGCCAAGACCTATCTTCTCTTCTCCCAATGACGAGGTGATGATCAACAA	1033	QY	700	ProGluGluMetAspLeuSerGlnGluGlnPro	710
QY	340	snLeuAlaTyrTyrAlaAlaMetLeuGlyGluGluHisThrArgSerIleGlyProArgG	360	Ds	2113	CCAGAGAGATGACCTCTCCAGAGAGAGCC	2145
Ds	1034	ATTTCGCTATTATTCAGCTATGCTTGGAGAAAGAACACACAGATCCATCGGCCCGCTG	1093	RESULT 7			
QY	360	uSerAlaLysGluTyrArgGlnArgSerLeuLeuGluLysGluLeuLeuPhePheAlaT	380	AA08489			
Ds	1094	AGAGTGCACAGAGTACCGACAGCGAAGCTTCTGGAAGAAAGAACTGCTTTCTTCGCTT	1153	ID	AA08489	standard; cDNA; 2577 BP	
QY	380	YrAspValPheGlyIleProPheValAspProAspSerTyrThrProGluGluValIleP	400	XX	AA08489;		
Ds	1154	ATGATGTTTTCGAATTCCTTTGATGATCCGATTCATGACTCCAGAGAGTATTC	1213	DT	09-AUG-2001	(first entry)	
QY	400	roLysArgLeuGlnGluLysGlnLysSerGluArgGluThrAlaValArgIleSerGlnG	420	DE	Human secreted protein-encoding gene 2 cDNA clone HUVF29, SEQ ID NO:12.		
Ds	1214	CCAGAGAGTTCAGAGAGAAACAGAAAGTCCAGAACCGGAAACAGCGTACGATCTCCAGG	1273	XX	Human; secreted protein; proliferative disorder; cancer; tumour;		
QY	420	luileGlyAsnLeuMetLysGluIleGluThrLeuValGluGluLysThrLysGluSerL	440	KW	fetal abnormality; developmental abnormality; haematopoietic disorder;		
Ds	1274	AGATTGGGACCTTATGAAGAAATCGAGACCTTGTGAGAGAGAGACCAAGGAGTAC	1333	KW	immune system disorder; AIDS; autoimmune disease; rheumatoid arthritis;		
QY	440	euAspValSerArgLeuThrArgGluGlyProLeuLeuTyrGluGlyIleSerLeuT	460	KW	inflammation; allergy; neurological disorder; Alzheimer's disease;		
Ds	1334	TGGATGTGAGCAGACTGACCCGGAAGGTGGCCCTCTGTATGAGGAGCATCATGTCTCA	1393	KW	Parkinson's disease; cognitive disorder; schizophrenia; asthma;		
QY	460	hrMetAsnSerLysLeuLeuAsnGlyTyr-GlnArgValValMetAspGlyValIleSer	479	KW	skin disorder; psoriasis; sepsis; diabetes; atherosclerosis;		
Ds	1394	CCATGAATCTCCAAATCTCTGAATGTTTC-CCAGCGGTGTGTGAGCGGTATCTCT	1452	KW	cardiovascular disorder; angiogenic disorder; kidney disorder;		
QY	480	AspHisGluCysGlnGlnGluLeuThrAsnValAlaAlaThrSerGlyAspGly	499	KW	gastrointestinal disorder; pregnancy-related disorder; gene therapy;		
Ds	1453	GACCAAGAGTGTGAGAGCTGACAGACTGACCAATGTGGCAGCAACCTCAGGAGATGGC	1512	KW	endocrine disorder; infection; wound healing; vulnery; cell culture;		
QY	500	TyrArgGlyGlnThrSerProHisThrProAsnGluLysPheTyrGlyValThrValPhe	519	KW	chemotaxis; food additive; binding partner identification; ss.		
Ds	1513	TACCGGGGTGACAGCTCCCACTACTCCCAATGAAAGTTCATGTGGTGTCTCTCTCTC	1572	XX	Homo sapiens.		
QY	520	LysAlaLeuLysLeuGlyGlnGlyLysValProLeuGlnSerAlaHisLeuTyrTyr	539	XX	Key	Location/Qualifiers	
Ds	1573	AAAGCCCTCAAGCTGGGGCAAGAGCAAGATTCCTCTGAGAGTGGCCCACTGTACTAC	1632	PH	CDS	24..1103	
				FT		/tag= a	
				FT		/product= "Human secreted protein precursor"	
				FT	sig_peptide	24..77	
				FT		/tag= b	
				FT	mat_peptide	78..1100	
				FT		/tag= c	
				FT		/product= "Mature human secreted protein"	
				XX			
				XX	WO200136432-A2.		
				XX	25-MAY-2001.		
				PD			

XX		15-NOV-2000; 2000WO-US031162.
PF		
XX		
XX		19-NOV-1999; 99US-0166415P.
PR		30-JUN-2000; 2000US-0215136P.
XX		
XX		(HUMA-) HUMAN GENOME SCI INC.
PA		
XX	Ruben SM, Komatsoulis GA, Baker KP, Young PE;	
PI	WPI; 2001-343793/36.	
DR	P-PSDB; AAE04200.	
XX		
XX	Isolated nucleic acid molecule encoding a human secreted protein is used	
PT	in preventing, treating or ameliorating a medical condition.	
XX		
XX	Claim 1; Page 403-404; 509pp; English.	
XX	AAD08488-AAD08529 represent cDNAs corresponding to 18 human secreted	
CC	protein genes, and AAE04199-AAE04239 represent the proteins they encode.	
CC	AAE04240-AAE04297 represent human secreted protein fragments or variants.	
CC	The secreted proteins and their genes are useful for preventing, treating,	
CC	or ameliorating medical conditions, e.g., by protein or gene therapy.	
CC	Pathological conditions can be diagnosed by determining the amount of the	
CC	new protein in a sample or by determining the presence of mutations in	
CC	the new genes. Specific uses are described for each of the 18 genes.	
CC	based on the tissues in which they are most highly expressed, and include	
CC	developing products for the diagnosis or treatment of proliferative	
CC	disorders, cancer, tumours, foetal and developmental abnormalities,	
CC	haematopoietic disorders, diseases of the immune system, AIDS, autoimmune	
CC	diseases (e.g., rheumatoid arthritis), inflammation, allergies,	
CC	neurological disorders (e.g., Alzheimer's disease, Parkinson's disease),	
CC	cognitive disorders, schizophrenia, asthma, skin disorders (e.g.,	
CC	psoriasis) sepsis, diabetes, atherosclerosis, cardiovascular disorders,	
CC	angiogenic disorders, kidney disorders, gastrointestinal disorders,	
CC	pregnancy-related disorders, endocrine disorders, and infections. The	
CC	proteins can also be used to aid wound healing and epithelial cell	
CC	proliferation, to prevent skin ageing due to sunburn, to maintain organs	
CC	before transplantation, for supporting cell culture of primary tissues,	
CC	to regenerate tissues, to identify their cognate ligands or binding	
CC	partners, and in chemotaxis, and can be used as a food additive or	
CC	preservative to modify storage properties. Antibodies specific for a	
CC	protein of the invention can be used in alleviating symptoms associated	
CC	with the disorders mentioned above, and in diagnostic immunoassays e.g.,	
CC	radioimmunoassay or enzyme linked immunosorbent assay (ELISA). The	
CC	present sequence represents a human secreted protein-encoding cDNA of the	
CC	invention	
XX		
SQ	Sequence 2577 BP; 610 A; 753 C; 711 G; 503 T; 0 U; 0 Other;	
Alignment Scores:		
Pred. No.:	0	Length: 2577
Score:	407.00	Matches: 733
Percent Similarity:	98.92%	Conservative: 0
Best Local Similarity:	98.92%	Mismatches: 3
Query Match:	55.30%	Indels: 8
DB:	4	Gaps: 0
US-10-045-815-4 (1-736) x AAD08489 (1-2577)		
Qy	1 MetAlaValArgAlaLeuLysLeuThrThrLeuLeuAlaValAlaAlaSer 20	
Dd	24 ATGCGGGTACGGCGGTGAAGCTGCTGACCACACTGCTGGCTGTCGGCGGTGCCTCC 83	
Qy	21 GlnAlaGluValGlnSerGluAlaGlyTrpGlyMetValThrProAspLeuLeuPheAla 40	
Dd	84 CAAGCCGAGTCCAGTCCGAGCAGATGGGGCATGGTGACGCCCTGATCTGCTCTTCGCC 143	
Qy	41 GluGlyThrAlaAlaTyrrAlaArgClyAspTrpProGlyValValLeuSerMetGluArg 60	
Dd	144 GAGGGACCGCAGCTACCGCGCGGGACTGCGCCGGTGGTCTTCCTGACATGGGAACGG 203	
Qy	61 AlaLeuArgSerArgAlaAlaLeuAlaLeuArgAlaLeuArgLeuArgCysArgThrGlnCysAla 80	

Db 1282 GGAGATTGGGAACCTTATGAGGAAATCGAGACCCCTTGTGAGAGAGAACCAAGGAGTC 1341  
Qy 439 rleuaspvalserargleuthrargglucglycylproleuthrargglucglylleSerLe 459  
Db 1342 ACTGGATGTGACGACACTGACCCGGGAAGGGTGGCCCCCTGCTGATGAGGCATCAGTCT 1401  
Qy 459 uThrMetAsnSerLysLeuLeuAsnGlyTyr-GlnArgValValMetAspGlyValIleS 479  
Db 1402 CACCATGAACTCCAACTCCTGAATGGTTC-CCAGCGGGTGGTATGAGCGCGTAATCT 1460  
Qy 479 erAspHisGluCysGlnGluLeuGlnArgLeuThrAsnValAlaAlaThrSerGlyAspG 499  
Db 1461 CTGACACGAGTGTGAGAGCTGCAGAGACTGACCAATGTGGCAGCAACCTCAGAGATG 1520  
Qy 499 lyTyrArgGlyGlnThrSerProHisThrProAsnGluLysPheTyrGlyValThrValP 519  
Db 1521 GCTACCGGGTGCAGACTCCACATCTCCCATCTCCATGAAAGTTCTATGGTGTCTACTG 1580  
Qy 519 heLysAlaLeuLysLeuGlyGlnGluGlyLysValProLeuGlnSerAlaHisLeuTyrT 539  
Db 1581 TCAAAGCCCTCAAGTGGCGCAAGAGGCAAGTTCTCTGCAGAGTGCCTCCTGACT 1640  
Qy 539 yrAsnValThrGluLysValArgArgGileMetGluSerTyrPheArgLeuAspThrProL 559  
Db 1641 ACACGTGACGAGAGAGTGGCGGCATCTGAGAGTCTCTCTCGCTCGATAGCGCCC 1700  
Qy 559 euTyrPheSerTyrSerHisLeuValCysArgThrAlaIleGluGluValGlnAlaGluA 579  
Db 1701 TCTACTTTTCTTACTCTCATCTGTGTGTCGCGCACTGCATCGAAGAGTCCAGGCAGAGA 1760  
Qy 579 rgLysAspAspSerHisProValHisValAspAsnCysIleLeuAsnAlaGluThrLeuV 599  
Db 1761 GGAAGGATGATAGTCATCCAGTCCACGTTGGACACTGTCATCTCTGAATGCCAGACCTCTG 1820  
Qy 599 aiCysValLysGluProProAlaTyrThrPheArgAspTyrSerAlaIleLeuTyrLeuA 619  
Db 1821 TGTGTGTCAAGAGAGCCCGCCGCTACCTTCGCGCACTCAGCGCCATCTTTTACCTAA 1880  
Qy 619 snGlyAspPheAspGlyGlyAsnPheTyrPheThrGluLeuAspAlaLysThrValThrA 639  
Db 1881 ATGGGGACTTCGATGGCGGAACTTTTATTTTCACTGAATCCGATGCCAGAGCCCTGACG 1940  
Qy 639 laGluValGlnProGlnCysGlyArgAlaValAlaGlyPheSerSerGlyThrGluAsnProH 659  
Db 1941 CAGAGTGCAGCTCAGTGTGGAGAGCGGTGGGATTTCTTTCAGGCACTGAAACCCAC 2000  
Qy 659 isGlyValLysAlaValThrArgGlyGlnArgCysAlaIleAlaLeuThrPheThrLeuA 679  
Db 2001 ATGGAGTGAAGCTGTCCAGCGGGCGACGCTGTGTCATCGCTCGCTGTGTTCACCTGG 2060  
Qy 679 spProArgHisSerGluArgAspArgValGlnAlaAspAspLeuValLysMetLeuPheS 699  
Db 2061 ACCCTCGACACAGCGAGCGGACAGGTCAGAGAGATGACTGCTGGTGAAGATGCTCTTCA 2120  
Qy 699 srProGluGluMetAspLeuSerGlnGluGlnProLeuAspAlaGlnGlnGlyProProG 719  
Db 2121 GCCCAGAAGAGATGACCTCTCCAGAGAGAGCGCCCTGGATGCCAGCAGCGGCCGCCCG 2180  
Qy 719 luproAlaGlnLysLeuSerGlySerGluSerLysProLysAspGluLeu 736  
Db 2181 AACCTGCAAGAGTCTCTCAGGCACTGAATCAGACCCCAAGATGAGCTA 2233

## RESULT 8

AAD08507

ID AAD08507 standard; cDNA; 2615 BP.

XX AAD08507;

AC AAD08507;

XX 09-AUG-2001 (first entry)

DE Human secreted protein-encoding gene 2 cDNA clone HUVFY29, SEQ ID NO:30.

XX Human; secreted protein; proliferative disorder; cancer; tumour;

foetal abnormality; developmental abnormality; haematopoietic disorder;  
immune system disorder; AIDS; autoimmune disease; rheumatoid arthritis;  
inflammation; allergy; neurological disorder; Alzheimer's disease;  
Parkinson's disease; cognitive disorder; schizophrenia; asthma;  
skin disorder; psoriasis; sepsis; diabetes; atherosclerosis;  
cardiovascular disorder; angiogenic disorder; kidney disorder;  
gastrointestinal disorder; pregnancy-related disorder; gene therapy;  
endocrine disorder; infection; wound healing; vulnerability; cell culture;  
chemotaxis; food additive; binding partner identification; ss.  
Homo sapiens.  
Key Location/Qualifiers  
CDS 24..545  
FT /tag= a  
FT /product= "Human secreted protein precursor"  
FT sig\_peptide 24..77  
FT /tag= b  
FT mat\_peptide 78..542  
FT /tag= c  
FT /product= "Mature human secreted protein"  
XX WO200136432-A2.  
XX 25-MAY-2001.  
XX 15-NOV-2000; 2000WO-US031162.  
XX 19-NOV-1999; 99US-0166415P.  
XX 30-JUN-2000; 2000US-0215136P.  
XX (HUMA-) HUMAN GENOME SCI INC.  
XX Ruben SM, Komatsoulis GA, Baker KP, Young PE;  
XX WPI; 2001-343793/36.  
XX P-PSDB; AAE04217.  
XX Isolated nucleic acid molecule encoding a human secreted protein is used  
XX in preventing, treating or ameliorating a medical condition.  
XX Claim 1; Page 416-417; 509pp; English.

AAD08488-AAD08529 represent cDNAs corresponding to 18 human secreted  
protein genes, and AAE04199-AAE04239 represent the proteins they encode.  
AAE04240-AAE04297 represent human secreted protein fragments or variants.  
The secreted proteins and their genes are useful for preventing, treating  
or ameliorating medical conditions, e.g., by protein or gene therapy.  
Pathological conditions can be diagnosed by determining the amount of the  
new protein in a sample or by determining the presence of mutations in  
the new genes. Specific uses are described for each of the 18 genes,  
based on the tissues in which they are most highly expressed, and include  
developing products for the diagnosis or treatment of proliferative  
disorders, cancer, tumours, foetal and developmental abnormalities  
haematopoietic disorders, diseases of the immune system, AIDS, autoimmune  
diseases (e.g., rheumatoid arthritis), inflammation, allergies,  
neurological disorders (e.g., Alzheimer's disease, Parkinson's disease),  
cognitive disorders, schizophrenia, asthma, skin disorders (e.g.,  
psoriasis), sepsis, diabetes, atherosclerosis, cardiovascular disorders,  
angiogenic disorders, kidney disorders, gastrointestinal disorders,  
pregnancy-related disorders, endocrine disorders, and infections. The  
proteins can also be used to aid wound healing and epithelial cell  
proliferation, to prevent skin ageing due to sunburn, to maintain organs  
before transplantation, for supporting cell culture of primary tissues,  
to regenerate tissues, to identify their cognate ligands or binding  
partners, and in chemotaxis, and can be used as a food additive or  
preservative to modify storage properties. Antibodies specific for a  
protein of the invention can be used in alleviating symptoms associated  
with the disorders mentioned above, and in diagnostic immunoassays e.g.,  
radioimmunoassay or enzyme linked immunosorbent assay (ELISA). The  
present sequence represents a human secreted protein-encoding cDNA of the  
invention

SQ Sequence 2615 BP; 616 A; 764 C; 717 G; 514 T; 0 U; 4 Other;

Alignment Scores:  
Pred. No.: 0 Length: 2615  
Score: 390.00 Matches: 731  
Percent Similarity: 98.92% Conservative: 0  
Best Local Similarity: 98.92% Mismatches: 5  
Query Match: 52.99% Indels: 8  
DB: 4 Gaps: 0

US-10-045-815-4 (1-736) x AAD08507 (1-2615)

QY	1	MetAlaValArgAlaLeuLysLeuThrThrLeuLeuAlaValValAlaAlaLaser	20
DB	24	ATGGCGGTAGCGCGTTGAAGCTGTGACACCACTGCTGGCTGTCTGGCGCTGCTCC	83
QY	21	GlnAlaGluValGluSerGluAlaGlyTrpGlyMetValThrProAspLeuPheAla	40
DB	84	CAAGCCGAGTTCGAGTCCGAGGCGAGGATGGGCGATGGTGGCGCTGATCTGCTTCGCG	143
QY	41	GluGlyThrAlaAlaTyrAlaArgGlyAspTrpProGlyValValLeuSerMetGluArg	60
DB	144	GAGGGGACCGCAGCTACGGCGCGGGGACTGGCGGGGTGGTCTGTGAGCATGGAAACGG	203
QY	61	AlaLeuArgSerArgAlaAlaLeuArgAlaLeuArgLeuArgCysArgThrGlnCysAla	80
DB	204	CGCTGGCTCCCGGGGAGCCCTCCGGCGCTTCGCTGGCTGGCGACCCAGTGTGCC	263
QY	81	AlaAspPheProTrpGluLeuAspProAspTrpSerProSerProAlaGlnAlaSerGly	100
DB	264	GCCGACTTCGCGTGGGAGCTGGACCCGACTGGTTCGCCACGCCGCGCCAGGCTCGGGC	323
QY	101	AlaGly-AlaLeuArgAspLeuSerPheGlyGlyLeuLeuArgArgAlaAlaCysLe	120
DB	324	CGCCG-CGCCCTGGCGGACCTGAGCTTCTCGGGGGCTTCTGGCTGGCGCTCGCT	382
QY	120	uArgArgCysLeuGlyProProAlaAlaHisSerLeuSerGluGluMetGluLeuGluPh	140
DB	383	CGCGCGTGGCTCGGGCGCGCGCGCGC-CACTCGCTCAGCGAAAGAGATGGAGTGGAGTT	441
QY	140	eArgGlyArgSerProTrpAsnTyrLeuGlnValAlaTyrPheLysIleAsnLysLeuG	160
DB	442	CCGCAAGCGGAC-CCCTACAACTACCTGCAGGTGGCTTCTCAAGATCAACAAAGTTGGA	500
QY	160	uLysAlaValAlaAlaAlaHisThrPhePheValGlyAsnProGluHisMetGluMetG	180
DB	501	GAAAGCTGTCTCGAGCACACACCTTCTTCGTGGCAATCTTGAGCACATGGAAATGCA	560
QY	180	nGlnAsnLeuAspTyrTyrGlnThrMetSerGlyValLysGluAlaAspPheLysAspLe	200
DB	561	GCAGAACCTAGACTATTACCAAAACCATGTCTGGAGTGAAGAGCGCGACTTCAAGGATCT	620
QY	200	uGluThrGlnProHisMetGlnGluPheArgLeuGlyValArgLeuTyrSerGluGluG	220
DB	621	TGAGACTCAACCCCATATGCAAGAAATTCGACTGGAGTGGGACTTACTCAGAGGAACA	680
QY	220	nProGlnGluAlaValProHisLeuGluAlaAlaLeuGlnGluTyrPheValAlaTyrG	240
DB	681	GCCAGGAAGCTGTGCCCCACCTAGAGCGCGCTGCAAGATACATTGTGGCGCTATGA	740
QY	240	uGluCysArgAlaLeuCysGluGlyProTyrAspTyrAspGlyTyrAsnTyrLeuGluT	260
DB	741	GGAGTGGCGTGGCTCTCGGAAGGCCCTATGACTACGATGGCTACAACTACCTTGAGTA	800
QY	260	rAsnAlaAspLeuPheGlnAlaIleThrAspHisTyrIleGlnValLeuAsnCysLysG	280
DB	801	CAACGCTGACCTTCCAGGGCCATCACAGATCATTTACATCCAGGTCTCTCAACTGTAA	860
QY	280	nAsnCysValThrGluLeuAlaSerHisProSerArgGluLysProPheGluAspPheLe	300
DB	861	GAACTGTGTACGGAGCTTGTCTCCACCCCAAGTCGAGAGAGCCCTTTGAAGACTTCCT	920
QY	300	uProSerHisTyrAsnTyrLeuGlnPheAlaTyrTyrAsnIleGlyAsnTyrThrGlnAl	320

DB	921	CCCATCGCATTAATATATCTGCAGTTTGCCTACTATAACATTGGGAATTATACACAGC	980
QY	320	aGly-GluCysAlaLysThrTyrLeuLeuPhePheProAsnAspGluValMetAsnGlnA	340
DB	981	TGT-TGAATGTGCCAAGACCTATCTTCTCTTCCCAATGACGAGGTGATGAAACAAA	1039
QY	340	snLeuAlaTyrTyrAlaAlaMetLeuGlyGluGluHisThrArgSerIleGlyProArgG	360
DB	1040	ATTGGCCCTATTATGACGTATGCTTGGAGAAGAACACACACAGATCCATCGCCCGCTG	1099
QY	360	luSerAlaLysGluTyrArgGlnArgSerLeuLeuGluLysGluLeuLeuPhePheAlaT	380
DB	1100	AGAGTGCACAGAGTACCGACAGCGAAGCTTACTGGAAAAAGAACTGCTTTCTTCGCTT	1159
QY	380	yrAspValPheGlyIleProPheValAspProAspSerTyrThrProGluGluValIleP	400
DB	1160	ATGATGTTTTGGAAATCCCTTTTGGGATCCGATTCATGACCTCCAGAAAGATGATTC	1219
QY	400	roLysArgLeuGlnGluLysGlnLysSerGluArgGluThrAlaValArgIleSerGlnG	420
DB	1220	CCAAGAGATTGCAAGAGAAAACAGAACTCAGAAACGGGAAACAGCCGTACGCATCTCC	1279
QY	420	luIleGlyAsnLeuMetLysGluIleGluThrLeuValGluGluLysThrLysGluSerL	440
DB	1280	AGATTGGGAACCTTATGAAGAAATCGAGACCCCTTGTGGAAGAGAAAGAACAGGAGT	1339
QY	440	euAspValSerArgLeuThrArgGluGlyGlyProLeuLeuTyrGluGlyIleSerLeuT	460
DB	1340	TGATGTGAGCAGACTGACCCGGGAAGGTGGCCCTCTGTATGAAGGCATCACTCTCA	1399
QY	460	hrMetAsnSerLysLeuLeuAsnGlyTyr-GlnArgValValMetAspGlyValIleSer	479
DB	1400	CCATGAACCTCCAACTCCTGAATGGTTC-CCAGCGGGTGGTGATGGAGCGGGAATCTCT	1458
QY	480	AspHisGluCysGlnGluLeuGlnArgLeuThrAsnValAlaAlaThr-SerGlyAspGly	499
DB	1459	GACACAGAGTGTGAGAGCTGCAGAGACTGACCACTATGGCAGCAACCTCAGAGATGCG	1518
QY	500	TyrArgGlyGlnThrSerProHisThrProAsnGluLysPheTyrGlyValThrValPhe	519
DB	1519	TACCGGGGTACAGCTCCCACTATCCCAATGAAAGTTCTATGGTGTCTACTGTCTTC	1578
QY	520	LysAlaLeuLysLeuGlyGlnGluLysValProLeuGlnSerAlaHisLeuTyrTyr	539
DB	1579	AAAGCCCTCAAGCTGGGGCAAGAGCAAGTTCTCTGACAGTGGCCACCTGACTACTAC	1638
QY	540	AsnValThrGluLysValArgArgIleMetGluSerTyrPheArgLeuAspThrProLeu	559
DB	1639	AACGTGACGAGAGAGGTGGCGGCATCATGGAGTCTTCTCCGCTGGATACGCCCTTC	1698
QY	560	TyrPheSerTyrSerHisLeuValCysArgThrAlaIleGluGluValGlnAlaGluArg	579
DB	1699	TACTTTTCTCTCTCATCTGGTGTGCCACTGCCATCGAAGAGGTCCAGGACAGAGG	1758
QY	580	LysAspAspSerHisProValHisValAspAsnCysIleLeuAsnAlaGluThrLeuVal	599
DB	1759	AGGATGATGATCATTCAGTCCAGTGGCAACTGCATCTCTGAATGCCGAGACCTCGTG	1818
QY	600	CysValLysGluProProAlaTyrThrPheArgAspTyrSerAlaIleLeuTyrLeuAsn	619
DB	1819	TGTGTCAAAAGAGCCCGACCTACACTTCCGCGACTACAGCGCCATCTTCTTACCTAAAT	1878
QY	620	GlyAspPheAspGlyGlyAsnPheTyrPheThrGluLeuAspAlaLysThrValThrAla	639
DB	1879	GGGGAATTCGATGGCGGAACCTTTTATTTCATGAACTGGATGCCAAGACCGGTGACGGA	1938
QY	640	GluValGlnProGlnCysGlyArgAlaValGlyPheSerSerGlyThrGluAsnProHis	659
DB	1939	GAGTGCAGCCTCAGTGTGGAAGAGCGGTGGGATCTCTTCCAGGCACTGAAACCCACAT	1998
QY	660	GlyValLysAlaValThrArgGlyGlnArgCysAlaIleAlaLeuThrPheThrLeuAsp	679

Db 1999 GGAGTGAAGGCTGTACACAGGGGGGAGCGCTGTGCATCGCCCTGTGGTTACCCCTGGAC 2058  
 QY 680 ProArgHisSerGluArgAspArgValGlnAlaAspAspLeuValLysMetLeuPheSer 699  
 Db 2059 CTTGACACAGCGAGCGGACAGGGTGCAGGAGATGACCTGGTGAAGATGCTTTCAGC 2118  
 QY 700 ProGluGluMetAspLeuSerGlnGluGlnProLeuAspAlaGlnGlnGlyProGlu 719  
 Db 2119 CCAGAAGAGATGAGCTCTCCAGGAGCGAGCCCTGGATGCCAGCAGGGCCCCCGGAA 2178  
 QY 720 ProAlaGlnGluSerLysSerGlySerGluSerLysProLysAspGluLeu 736  
 Db 2179 CCTGCACAGAGTCTCTCAGGAGTGAATCGAAGCCCAAGGATGAGCTA 2229

RESULT 9

ABK94910  
 ID ABK94910 standard; cDNA; 2753 BP.  
 XX  
 AC ABK94910;

XX 30-AUG-2002 (first entry)

XX Human novel polynucleotide #21.

XX Human; gene; ss; inflammatory condition; shock; sepsis; immune response;  
 KW cancer; wound healing; central nervous system disease; haematopoiesis;  
 KW peripheral nervous system disease; amyotrophic lateral sclerosis; tendon;  
 KW myeloid cell disorder; lymphoid cell disorder; platelet disorder; bone;  
 KW cartilage; ligament; nerve tissue; ulcer; osteoporosis; osteoarthritis;  
 KW bone degenerative disorder; periodontal disease; reperfusion injury;  
 KW lung fibrosis; liver fibrosis; autoimmune disorder; bacterial infection;  
 KW allergic condition; thrombolytic; thrombosis; coagulation disorder;  
 KW fungal infection.

XX Homo sapiens.

XX WO200244340-A2.

XX 06-JUN-2002.

XX 30-NOV-2001; 2001WO-US047004.

XX 30-NOV-2000; 2000US-00728952.

XX (HYSE-) HYSEQ INC.

XX Tang YT, Goodrich RW, Liu C, Zhou P, Asundi V, Wang J, Wang D;  
 XX Yamazaki V, Ujwal ML, Drmanac RT;  
 XX P-PSDB; ABG66686.

XX WPI; 2002-508509/54.

XX P-PSDB; ABG66686.

PT Novel nucleic acids and polypeptides for diagnosis, treatment of  
 PT inflammatory, autoimmune, nervous system, myeloid or lymphoid cell  
 PT disorders, cancer and promoting wound healing.

XX Claim 1; Page 369-373; 672pp; English.

XX The invention relates to human novel polynucleotides and associated  
 CC polypeptides. The polynucleotides and polypeptides are useful for  
 CC treating inflammatory conditions such as arthritis, nephritis, Crohn's  
 CC disease, ischaemia-reperfusion injury, shock, sepsis, immune responses  
 CC and cancer and for promoting wound healing. The sequences are used to  
 CC induce the proliferation of neural cells and regeneration of nerve and  
 CC brain tissue, and are useful for the treatment of central and peripheral  
 CC nervous system diseases and neuropathies, such as Alzheimer's disease,  
 CC Parkinson's disease, Huntington's disease and amyotrophic lateral  
 CC sclerosis. The sequences are involved in chemotactic or chemokinetic  
 CC activity, regulation of haematopoiesis, treatment of myeloid or lymphoid  
 CC cell disorders and platelet disorders such as thrombocytopenia,  
 CC regeneration of bone, cartilage, tendon, ligament and/or nerve tissue  
 CC growth, tissue repair, healing of burns, incisions, ulcers, treatment of  
 CC osteoporosis, osteoarthritis, bone degenerative disorders and periodontal

CC disease. The sequences of the invention are also useful for gut  
 CC protection or regeneration and treatment of lung or liver fibrosis,  
 CC reperfusion injury in various tissues, immune deficiencies and disorders  
 CC including severe combined immunodeficiency (SCID), bacterial or fungal  
 CC infections, autoimmune disorders e.g. multiple sclerosis and myasthenia  
 CC gravis, allergic conditions such as asthma, thrombolytic or thrombotic  
 CC and coagulation disorders. Sequences ABK94890-ABK94982 represent human  
 CC novel polynucleotides of the invention  
 XX

SQ Sequence 2753 BP; 632 A; 802 C; 784 G; 535 T; 0 U; 0 Other;

Alignment Scores:

Pred. No.: 0 Length: 2753  
 Score: 380.00 Matches: 580  
 Percent Similarity: 99.32% Conservative: 0  
 Best Local Similarity: 99.32% Mismatches: 2  
 Query Match: 51.63% Indels: 4  
 DB: 6 Gaps: 0

US-10-045-815-4 (1-736) x ABK94910 (1-2753)

QY 155 LysIleAsnLysLeuGluLysAlaValAlaAlaHisThrPhePheValGlyAsnPro 174  
 Db 673 AAGATCAACAAAGTTGGAGAAAGCTGTGTGTCAGCACACACCTCTTCTGTGGCAATCCT 732  
 QY 175 GluHisMetGluMetGlnGlnAsnLeuAspTyrTyrGlnThrMetSerGlyValLysGlu 194  
 Db 733 GAGCATGTGAATTCGAGCAGCACTAGCTATTACCAACCATCTCTGGAGTGAAGAG 792  
 QY 195 AlaAspPheLysAspLeuGluThrGlnProHisMetGlnGluPheArgLeuGlyValArg 214  
 Db 793 GCGCATCTTCAGGATCTTGAGACTCAACCCCATATGCAAGAAATTCGACTGGAGTGGCA 852  
 QY 215 LeuTyrSerGluGluGlnProGlnGluAlaValProHisLeuGluAlaAlaLeuGlnGlu 234  
 Db 853 CTCTACTCAGAGAAACAGCCAGGAAGCTGTGCCACCTAGAGCGGCGCTGCAAGAA 912  
 QY 235 TyrPheValAlaTyrGluGluCysArgAlaLeuCysGluGlyProTyrAspTyrAspGly 254  
 Db 913 TACTTTGTGGCTATGAGAGTGGCTGCCCTCTCGAAGGGCCCTATGACTACGATGCG 972  
 QY 255 TyrAsnTyrLeuGluTyrAsnAlaAspLeuPheGlnAlaIleThrAspHisTyrIleGln 274  
 Db 973 TACAACCTACCTTGATGATCAACGCTGACCTCTCCAGGCCATCACAGATCATTTACATCCAG 1032  
 QY 275 ValLeuAsnCysLysGlnAsnCysValThrGluLeuAlaSerHisProSerArgGluLys 294  
 Db 1033 GTCTCAACTGTATAGCAGCACTGTGTACAGGAGCTTGCTCCACCCCAAGTCGAGAGAG 1092  
 QY 295 ProPheGluAspPheLeuProSerHisTyrAsnTyrLeuGlnPheAlaTyrTyrAsnIle 314  
 Db 1093 CCGTTTGAAGACTTCTCCCATTCGATTAATATATCTGCAGTTTGGCTACTATAACATT 1152  
 QY 315 GlyAsnTyrThrGlnAlaGly-GluCysAlaLysThrTyrLeuLeuPhePheProAsnAs 334  
 Db 1153 GGGATTTATACACAGGCTGT-TGAATGTGCCAAGACCTATCTTCTCTTCCCAATGA 1211  
 QY 334 pGluValMetAsnGlnAsnLeuAlaTyrTyrAlaAlaMetLeuGlyGluGluHisThrAr 354  
 Db 1212 CGAGGTGATGAACCAAAATTTGGCTATTATGACGCTATGCTTGGAGAGAACACACAG 1271  
 QY 354 gSerIleGlyProArgGluSerAlaLysGluTyrArgGlnArgSerLeuLeuGlyGlu 374  
 Db 1272 ATCCATCGCCGCCCTGAGAGTGCAGAGGAGTCCGACAGCAAGGCTACTTGGAAAAAGA 1331  
 QY 374 uLeuLeuPhePheAlaTyrAspValPheGlyIleProPheValAspProAspSerTrpTh 394  
 Db 1332 ACTGCTTTCTTCGCTTATGATGTTTTTGGAAATTCCTTTTGGATCCGGATTCATGGAC 1391  
 QY 394 rProGluGluValIleProLysArgLeuGlnGluLysGlnLysSerGluArgGluThrAl 414  
 Db 1392 TCCAGAGAGAGTGAATCCCAAGAGATTGCAAGAGAAACAGAGATCGGAAAAAGCAGC 1451

Qy	414	aValArgIleSerGlnGluLeuGlyAsnLeuMetLysGluLeuThrLeuValGluG	434
Db	1452	CGTACGCATCTCCAGAGATGGGAACCTTATGAAGAAATCGACACCTTGTGAAGA	1511
Qy	434	uLysThrLysGluSerLeuAspValSerArgLeuThrArgGluGlyGlyProLeuLeuTy	454
Db	1512	GAAGACCAAGAGTCACTGATGTGACGACACTGACCCGGAAGTGGCCCTGTCTGTA	1571
Qy	454	rGluGlyIleSerLeuThrMetAsnSerLysLeuLeuAsnGlyTyr-GlnArgValValM	474
Db	1572	TGAAGGATCAGTCTCAACATGAATCAACCTCACTGAATGGTTC-CCAGCGGGTGTGA	1630
Qy	474	eAspGlyValIleSerAspHisGluCysGlnGluLeuGlnArgLeuThrAsnValAlaA	494
Db	1631	TGGACGGCGTAATCTTGACACAGAGTGTGAGAGCTGCAGAGACTGACCAATGTGCAG	1690
Qy	494	laThrSerGlyAspGlyTyrArgGlyGlnThrSerProHisThrProAsnGluLysPheT	514
Db	1691	CAACCTCAGGAGATGGCTACGGGGTCAGACCTCCCCACATACTCCCAATGAAGAATCT	1750
Qy	514	yrGlyValThrValPheIysAlaLeuLysLeuGlyGlnGluGlyLysValProLeuGlnS	534
Db	1751	ATGGTGTCACGTCTTCAAGCCCTCAAGCTGGGGCAAGAGCAAGTTCTCTGCAGA	1810
Qy	534	erAlaHisLeuTyrTyrAsnValThrGluLysValArgArgIleMetGluSerTyrPheA	554
Db	1811	GTGCCCACTGTACTACACGTGACGGAAGGTGGCGGCATCATGGAGTCTCTACTTC	1870
Qy	554	rgLeuAspThrProLeuTyrPheSerTyrSerHisLeuValCysArgThrAlaIleGluG	574
Db	1871	GCCTGGATACGCCCTCTACTTTCTCTACTCTCATCTGGTGGCGGACATGCCATCGAAG	1930
Qy	574	luValGlnAlaGluArgLysAspSerHisProValHisValAspAsnCysIleLeuA	594
Db	1931	AGTCTCAGGACAGAGAGATGATGATCATCTCAGTCCAGTCCAGTCACTGCATCTCGA	1990
Qy	594	snAlaGluThrLeuValCysValLysGluProProAlaTyrThrPheArgAspTyrSerA	614
Db	1991	ATGCGGAGACCTCGTGTGTCTCAAGAGCGCCCGACCTACACCTTCCGGGACTACAGCG	2050
Qy	614	laIleLeuTyrLeuAsnGlyAspPheAspGlyGlyAsnPheTyrPheThrGluLeuAspA	634
Db	2051	CCATCTTTACTTAAATGGGACCTTCGATGGCGGAACTTTTATTCACGTGAATGGATG	2110
Qy	634	laLysThrValThrAlaGluValGlnProGlnCysGlyArgAlaValGlyPheSerSerG	654
Db	2111	CCAAAGCCGTACGCGCAGAGTGCAGCTCAGTGTGGAAGAGCGGTGGGATTTCTTTCAG	2170
Qy	654	lyThrGluAsnProHisGlyValLysAlaValThrArgGlyGlnArgCysAlaIleAlaL	674
Db	2171	GCACTGAAACCCACATGGAGTGAAGGCTGTACACAGGGGGGACGCTGTGCCATCGCCC	2230
Qy	674	euTppPheThrLeuAspProArgHisSerGluArgAspArgValGlnAlaAspAspLeuV	694
Db	2231	TGTGTTTCACTGGACCTCGACACAGCAGCGGACAGGGTGCAGGCAGATGACCTGG	2290
Qy	694	alLysMetLeuPheSerProGluGluMetAspLeuSerGlnGluGlnProLeuAspAlaG	714
Db	2291	TGAAGATGCTCTTCAGCCCGAAGAGATGAGACCTCTCCAGAGACAGCCCTGGATGCC	2350
Qy	714	lnGlnGlyProProGluProAlaGlnGluSerLeuSerGlySerGluSerLysProLysA	734
Db	2351	AGCAGGGCCCCCGAACCTGCACAGAGTCTCTCTCAGGGCAGTGAATCGNAGCCCAAGG	2410
Qy	734	spGluLeu	736
Db	2411	ATGAGCTA	2418
XX	RESULT 10		
XX	ABK94933		
XX	ID	ABK94933	standard; cDNA; 2753 BP.
XX	AC	ABK94933;	

XX	30-AUG-2002	(first entry)	
DT	Human novel polynucleotide #44.		
DE	Human; gene; ss; inflammatory condition; shock; sepsis; immune response; cancer; wound healing; central nervous system disease; haematopoiesis; peripheral nervous system disease; amyotrophic lateral sclerosis; tendon; myeloid cell disorder; lymphoid cell disorder; platelet disorder; bone; cartilage; ligament; nerve tissue; ulcer; osteoporosis; osteoarthritis; bone degenerative disorder; periodontal disease; reperfusion injury; lung fibrosis; liver fibrosis; autoimmune disorder; bacterial infection; allergic condition; thrombolytic; thrombosis; coagulation disorder; fungal infection.		
XX	Homo sapiens.		
OS	WO200244340-A2.		
XX	06-JUN-2002.		
PD	30-NOV-2001; 2001WO-US047004.		
XX	30-NOV-2000; 2000US-00728952.		
XX	(HYSE-) HYSEQ INC.		
XX	Tang YT, Goodrich RW, Liu C, Zhou P, Asundi V, Wang J, Wang D; Yamazaki V, Ujwal ML, Drmanac RT;		
PI	WPI; 2002-508509/54.		
XX	P-PSDB; ABG66709.		
DR	Novel nucleic acids and polypeptides for diagnosis, treatment of inflammatory, autoimmune, nervous system, myeloid or lymphoid cell disorders, cancer and promoting wound healing.		
PT	Claim 1; Page 441-444; 672pp; English.		
XX	The invention relates to human novel polynucleotides and associated polypeptides. The polynucleotides and polypeptides are useful for treating inflammatory conditions such as arthritis, nephritis, Crohn's disease, ischaemia-reperfusion injury, shock, sepsis, immune responses and cancer and for promoting wound healing. The sequences are used to induce the proliferation of neural cells and regeneration of nerve and brain tissue, and are useful for the treatment of central and peripheral nervous system diseases and neuropathies, such as Alzheimer's disease, Parkinson's disease, Huntington's disease and amyotrophic lateral sclerosis. The sequences are involved in chemotactic or chemokinetic activity, regulation of haematopoiesis, treatment of myeloid or lymphoid cell disorders and platelet disorders such as thrombocytopenia, tissue regeneration of bone, cartilage, tendon, ligament and/or nerve tissue growth, tissue repair, healing of burns, incisions, ulcers, treatment of osteoporosis, osteoarthritis, bone degenerative disorders and periodontal disease. The sequences of the invention are also useful for gut protection or regeneration and treatment of lung or liver fibrosis, reperfusion injury in various tissues, immune deficiencies and disorders including severe combined immunodeficiency (SCID), bacterial or fungal infections, autoimmune disorders e.g. multiple sclerosis and myasthenia gravis, allergic conditions such as asthma, thrombolytic or thrombosis and coagulation disorders. Sequences ABK94890-ABK94982 represent human novel polynucleotides of the invention		
XX	Sequence 2753 BP; 632 A; 802 C; 784 G; 535 T; 0 U; 0 Other;		
SQ	Alignment Scores:		
	Pred. No.:	0	Length: 2753
	Score:	380.00	Matches: 580
	Percent Similarity:	99.32%	Conservative: 0
	Best Local Similarity:	99.32%	Mismatches: 2
	Query Match:	51.63%	Indels: 4
	DB:	6	Gaps: 0



US-10-045-815-4 (1-736) x ABK94933 (1-2753)

QY 155 LysTlleAsnLysLeuGluLysAlaValAlaAlaAlaHisThrPhePheValGlyAsnPro 174  
Db 673 AAGATCAACAAGATTGGAGAAAGCTGTTCCTGAGCACAACCTTCTTCGTGGCAATCCT 732  
QY 175 GluHisMetGluMetGlnGlnAsnLeuAspTyrTyrGlnThrMetSerGlyValLysGlu 194  
Db 733 GAGCACAATGGAAATGACAGAGAACTAGACTATTACCAACCAATGTCTGGAGTGAAGAG 792  
QY 195 AlaAspPheLysAspLeuGluThrGlnProHisMetGlnGluPheArgLeuGlyValArg 214  
Db 793 GCCGACTTCAAGGATCTTGGAGACTCAACCCATATGCAAGAAATTTGACATGGGAGTGGC 852  
QY 215 LeuTyrSerGluGluGlnProGlnGluAlaValProHisLeuGluAlaAlaLeuGlnGlu 234  
Db 853 CTCTACTCAGAGAAACACCCACAGAGACTGTGCCACCTAGAGGGCGCTGCAAGAA 912  
QY 235 TyrPheValAlaTyrGluGluCysArgAlaLeuCysGluGlyProTyrAspTyrAspGly 254  
Db 913 TACTTTGTGGCTATGAGGAGTGGCGCTCTCGGAAGGGCCCTATGACTACGATGGC 972  
QY 255 TyrAsnTyrLeuGluTyrAsnAlaAspLeuPheGlnAlaThrAspHisTyrIleGln 274  
Db 973 TACAACCTACTTGAGTAGCAACCGCTGACCTCTTCAGGCCATCACAGATCATTTACATCCAG 1032  
QY 275 ValLeuAsnCysLysGlnAsnCysValThrGluLeuAlaSerHisProSerArgGluLys 294  
Db 1033 GTCCTCACTGAGCAGAACTGTGTACGAGGCTTGTCTCCCAACCAAGTCGAGAGAAG 1092  
QY 295 ProPheGluAspPheLeuProSerHisTyrAsnTyrIleGlnPheAlaTyrTyrAsnIle 314  
Db 1093 CCTTTTGAAGACTTCTCTCCCATCGCAATATATATCTGCAAGTTTGGCTACTATAACATT 1152  
QY 315 GlyAsnTyrThrGlnAlaGly-GluCysAlaLysThrTyrLeuLeuPhePheProAsnAs 334  
Db 1153 GGGAAATATACAGAGCTGT-TGAATGTGCCAAGACCTATCTTCTCTTCCCAATGA 1311  
QY 334 pGluValMetAsnGlnAsnLeuAlaTyrTyrAlaAlaMetLeuGlyGluGluHisThrAr 354  
Db 1212 CGAGGTGATGAACCAAAATTTGGCCTATTATGAGCTATGCTTGCAGAAACACACACAG 1271  
QY 354 sSerIleGlyProArgGluSerAlaLysGluTyrArgGlnArgSerLeuLeuGluLysG 374  
Db 1272 ATCCATCGCCCCCGTGAGGTGCCAGGAGTACCGACAGAGCTACTCTGGAAGAAAGA 1331  
QY 374 uLeuLeuPhePheAlaTyrAspValPheGlyIleProPheValAspProAspSerTrpTh 394  
Db 1332 ACTGCTTTCTTCGCTTATGATGTTTTTGGAAATTCCTTTGTGGATCCGGATTCATGGAC 1391  
QY 394 rProGluGluValIleProLysArgLeuGlnGluLysGlnLysSerGluArgGluThrAl 414  
Db 1392 TCCAGAGAGAGTGAATCCAGAGATTTGCCAGACTTATGAGGAAATCGAGACCTCTTGGAGA 1451  
QY 414 aValArgIleSerGlnGluIleGlyAsnLeuMetLysGluIleGluThrLeuValGluG 434  
Db 1452 CGTACGATCTCCAGAGAGTGGAGACTTATGAGGAAATCGAGACCTCTTGGAGA 1511  
QY 434 uLysThrLysGluSerLeuAspValSerArgLeuThrArgGluGlyProLeuLeuTyr 454  
Db 1512 GAAGACCAAGGAGTCACTGGATGTGAGCAGACTGACCCGGGAAAGGTGGCCCCCTGTCTGA 1571  
QY 454 rGluGlyIleSerLeuThrMetAsnSerLysLeuLeuAsnGlyTyr-GlnArgValValM 474  
Db 1572 TGAAGGATCATGTCTACCATGAACTCCAACTCTGAAATGGTTC-CCAGCGGCTGGTGA 1630  
QY 474 etAspGlyValIleSerAspHisGluCysGlnGluLeuGlnArgLeuThrAsnValAla 494  
Db 1631 TGGACGGCGTAATCTCTGACCACAGGTGTGAGGAGCTGCAGAGACTGACCAATGTGGCAG 1690  
QY 494 laThrSerGlyAspGlyTyrArgGlyGlnThrSerProHisThrProAsnGluLysPheT 514  
Db 1691 CAACCTCAGGAGATGGCTACCGGGGTGAGACCTCCCACTCCCAATGAAAGATTCT 1750

QY 514 YrClyValThrValPheLysAlaLeuLysLeuGlyGlnGluGlyLysValProLeuGlnS 534  
Db 1751 ATGGTGTCACTGTCTTCAAAAGCCCTCAAGCTGGGCAAGAGCGAAAGTTCTCTCTGCAGA 1810  
QY 534 exAlaHisLeuTyrTyrAsnValThrGluLysValArgArgIleMetGluSerTyrPheA 554  
Db 1811 GTGCCCACTGTACTACAACGTGACGGAGAAGGTGGGGCATCATGAGTCTTACTTCC 1870  
QY 554 rGluAspThrProLeuTyrPheSerTyrSerHisLeuValCysArgThrAlaIleGluG 574  
Db 1871 GCCTGATACGCCCTCTACTTTTCTACTCTCATCTGGTGTGGCCACTGTCATCGAAG 1930  
QY 574 luValGlnAlaGluArgLysAspAspSerHisProValHisValAspAsnCysIleLeuA 594  
Db 1931 AGGTCCAGCAGAGAGAGAGAGATGATGATCATCTCAGTCCAGCTGGACAACATGCTCTGA 1990  
QY 594 snAlaGluThrLeuValCysValLysGluProProAlaTyrThrPheArgAspTyrSerA 614  
Db 1991 ATGCCGAGACCTCTGTGTGTGTCAAAGAGCCCCAGCCCTACACCTTCCGCACTACAGCG 2050  
QY 614 laIleLeuTyrLeuAsnGlyAspPheAspGlyGlyAsnPheTyrPheThrGluLeuAspA 634  
Db 2051 CCATCTTTTACTAAATGGGACTTTCGATGGCGGAAACTTTTATTCTACTGAACCTGGATG 2110  
QY 634 laLysThrValThrAlaGluValGlnProGlnCysGlyArgAlaValGlyPheSerSerG 654  
Db 2111 CCAAGACCGTGAGCGGAGAGGTGCGAGCTCAGTGTGGAAGAGCGGTGGGATTCCTTTCAG 2170  
QY 654 lyThrGluAsnProHisGlyValLysAlaValThrArgGlyGlnArgCysAlaIleAlaL 674  
Db 2171 GCATGAAACCCACATGAGTGAAGGTGTCCACGAGGGGCGAGCGCTGTGCCATCGCCC 2230  
QY 674 euTyrPheThrLeuAspProArgHisSerGluArgAspArgValGlnAlaAspAspLeuV 694  
Db 2231 TGTGTTTCACTTGGACCTCTGCACACAGCGAGCGGACAGGCTGCAGGACGATGACCTGG 2290  
QY 694 aLysMetLeuPheSerProGluGluMetAspLeuSerGlnGluGlnProLeuAspAlaG 714  
Db 2291 TGAAGATGCTCTTTCAGCCAGAGAGATGAGCTCTCCAGAGAGCAGCCCTGGATGCC 2350  
QY 714 lngGlnGlyProGluProAlaGlnGluSerLeuSerGlySerGluSerLysProLysA 734  
Db 2351 AGCAGGGCCCCCCCCAAGCTGCTCTCTCAGGAGTGTCTCTCAGGAGTGTGAGGCCAAGG 2410  
QY 734 spGluLeu 736  
Db 2411 ATGAGCTA 2418

## RESULT 11

AAC64724

ID AAC64724 standard; cDNA; 2829 BP.

XX AAC64724;

XX 27-FEB-2001 (first entry)

XX Human tumour suppressor Gros1-L encoding cDNA SEQ ID NO:1.

XX Tumour suppressor; Gros1-L; Gros1-S; cell proliferation; regulation;  
XX cancer; cytostatic; gene therapy; ss.  
XX Homo sapiens.  
XX WO200065047-A1.  
XX 02-NOV-2000.  
XX 26-APR-2000; 2000WO-JP002731.  
XX 26-APR-1999; 95JP-00118806.  
XX (CHUG-) CHUGAI RES INST MOLECULAR MEDICINE INC.



XX Wadhwa R, Sugihara T, Yoshida A;  
PI WPI: 2000-687340/67.  
DR P-PSDB; AAB36391.  
XX Full-length tumor suppressor genes encoding Gros1-L, Gros1-S and mouse  
PT homologs participating in regulation of cell proliferation, useful in  
PT development of preventives and remedies of cancer.  
XX Claim 1; Page 61-66; 114pp; Japanese.  
XX The present sequence encodes the human tumour suppressor designated Gros1  
CC -L. Gros1-L and Gros1-S have cytosolic activity and can be used in gene  
CC therapy. Gros1-L and Gros1-S genes are useful in the development of drugs  
CC used to treat and prevent cancer  
XX Sequence 2829 BP; 659 A; 808 C; 801 G; 561 T; 0 U; 0 Other;  
SQ  
Alignment Scores:  
Pred. No.: 0 Length: 2829  
Score: 360.00 Matches: 360  
Percent Similarity: 100.00% Conservative: 0  
Best Local Similarity: 100.00% Mismatches: 0  
Query Match: 48.91% Indels: 0  
DB: 3 Gaps: 0  
US-10-045-815-4 (1-736) x AAC64724 (1-2829)  
QY 1 MetAlaValArgAlaLeuLysLeuLeuThrThrLeuLeuAlaValValAlaAlaLaser 20  
DB 52 ATGGCGGTACGCGCTTGAAGCTGTGACACACTGCTGGCTGTGGCGCGCTGCTGC 111  
QY 21 GluAlaGluValGluSerGluAlaGlyTTPGlyMetValThrProAspLeuLeuPheAla 40  
DB 112 CAAGCCGAGTCTGAGTCCGAGGAGGATGGGGATGGTGTGACGCTGATCTGCTTCG 171  
QY 41 GluGlyThrAlaAlaValAlaArgGlyVasPTProGlyValValLeuSerMetGluArg 60  
DB 172 GAGGGACCCGAGCTACGCGCGCGGAGCTGGCCGGGGTGTCTCTGAGCATGGAAAG 231  
QY 61 AlaLeuArgSerArgAlaAlaLeuArgAlaLeuArgAlaLeuArgCysArgThrGlnCysAla 80  
DB 232 GCGTGGCTCCGGGCGAGCCCTCCGCGCCCTTCGCTGGCTGGCTGGCCAGCCAGTGTGC 291  
QY 81 AlaAspPheProTTPGluLeuAspProAspTTPSerProAlaGlnAlaSerGly 100  
DB 292 GCCGACTTCCCGTGGAGCTGGACCCGACTGTGTCCCGAGCCGCGCCAGGCTCGGGC 351  
QY 101 AlaGlyAlaLeuArgAspLeuSerPhePheGlyLeuLeuArgArgAlaAlaCysLeu 120  
DB 352 GCGCGCCCTCGCGGACCTGAGCTTCTTCGGGGGCTTCTGCGTGGCTGGCTGGCTG 411  
QY 121 ArgArgCysLeuGlyProProAlaAlaHisSerLeuSerGluLeuMetGluLeuGluPhe 140  
DB 412 CGCCGCTGCTCGGGCGCGCGCGCCGCTGCTGCTGAGGAGAGATGGAGCTGGAGTTC 471  
QY 141 ArgLysArgSerProTyrAsnTyrLeuGlnValAlaTyrPheLysIleAsnLysLeuGlu 160  
DB 472 CGCAAGCGGAGCCCTACAACTACTCTGAGGTGGCTTCAAGATCAACAAGATTGGAG 531  
QY 161 LysAlaValAlaAlaAlaHisThrPhePheValGlyAsnProGluHisMetGluMetGln 180  
DB 532 AAAGCTGTGTGTCGAGCACACACCTTCTTCTGGGCAATCTCTGAGCAATGGAAATG 591  
QY 181 GlnAsnLeuAspTyrTyrGlnThrMetSerGlyValLysGluAlaAspPheLysAspLeu 200  
DB 592 CAGAACCTAGACTATTACCAACCACTCTGGAGTGAGGAGCCGACTTCAAGATCTT 651  
QY 201 GluThrGlnProHisMetGlnGluPheArgLeuGlyValArgLeuTyrSerGluGluGln 220  
DB 652 GAGACTCAACCCCATATGCAAGATTTTCGACTGGAGTGGGACTCTACTCAGAGGAACAG 711

QY 221 ProGlnGluAlaValProHisLeuGluAlaAlaLeuGlnGluTyrPheValAlaTyrGlu 240  
DB 712 CCACAGGAAGCTGTGGCCACCTAGAGCGCGCTGCAAGAAATACTTTGTGGCCCTATGAG 771  
QY 241 GluCysArgAlaLeuLeuGluGlyProTyrAspTyrAspGlyTyrAsnTyrLeuGluTyr 260  
DB 772 GAGTCCCGTGGCTTGGAGAGGCCCTATGACTACGATGAGTGTGCTACAACTACTCTGAGTAC 831  
QY 261 AsnAlaAspLeuPheGlnAlaIleThrAspHisTyrIleGlnValLeuAsnCysLysGln 280  
DB 832 AACGCTGACCTTCTCCAGGCCATCACAGATCATACATCCAGGCTCCTCAACTGTAAAGCAG 891  
QY 281 AsnCysValThrGluLeuAlaSerHisProSerArgGluLysProPheGluAspPheLeu 300  
DB 892 AACTGTGTACAGGAGCTTGTTCGCCCAAGTCGAGAGAGCCCTTTGAAAGACTTCTTC 951  
QY 301 ProSerHisTyrAsnTyrLeuGlnPheAlaTyrTyrAsnIleGlyAsnTyrThrGlnAla 320  
DB 952 CCATCGCATTAATTAATTAATTCAGTTTGCCTTACTATAACATTGGGATTTATACAAAGCT 1011  
QY 321 GlyGluCysAlaLysThrTyrLeuLeuPhePheProAsnAspGluValMetAsnGlnAsn 340  
DB 1012 GGTGAATGTGCAAGACCTATCTTCTTCTTCCCAATGACGAGGTGATGACCAAAAT 1071  
QY 341 LeuAlaTyrTyrAlaAlaMetLeuGlyGluGluHisThrArgSerIleGlyProArgGlu 360  
DB 1072 TTGGCCTATTATGACGCTATGCTTGGAGAGACACACACAGATCCATCGGCCCGCTGAG 1131  
RESULT 12  
AAH16288  
ID AAH16288 standard; cDNA; 2993 BP.  
XX AAH16288;  
XX 26-JUN-2001 (first entry)  
DE Human cDNA sequence SEQ ID NO:15159.  
XX Human; primer; detection; diagnosis; antisense therapy; gene therapy; ss.  
XX Homo sapiens.  
XX EP1074617-A2.  
XX 07-FEB-2001.  
XX 28-JUL-2000; 2000EP-00116126.  
XX 29-JUL-1999; 99JP-00248036.  
XX 27-AUG-1999; 99JP-00300253.  
XX 11-JAN-2000; 2000JP-00118776.  
XX 02-MAY-2000; 2000JP-00183767.  
XX 09-JUN-2000; 2000JP-00241899.  
XX (HELI-) HELIX RES INST.  
XX Ota T, Isogai T, Nishikawa T, Hayashi K, Saito K, Yamamoto J;  
PI Ishii S, Sugiyama T, Wakamatsu A, Nagai K, Otsuki T;  
XX WPI: 2001-318749/34.  
XX Primer sets for synthesizing polynucleotides, particularly the 5602 full-  
PT length cDNAs defined in the specification, and for the detection and/or  
PT diagnosis of the abnormality of the proteins encoded by the full-length  
PT cDNAs.  
XX Claim 8; SEQ ID NO 15159; 2537pp + Sequence Listing; English.  
XX The present invention describes primer sets for synthesizing 5602 full-  
CC length cDNAs defined in the specification. Where a primer set comprises:  
CC (a) an oligo-dr primer and an oligonucleotide complementary to the  
CC complementary strand of a polynucleotide which comprises one of the 5602  
CC nucleotide sequences defined in the specification, where the

oligonucleotide comprises at least 15 nucleotides; or (b) a combination of an oligonucleotide comprising a sequence complementary to the complementary strand of a polynucleotide which comprises a 5'-end sequence and an oligonucleotide comprising a sequence complementary to a polynucleotide which comprises a 3'-end sequence, where the oligonucleotide comprises at least 15 nucleotides and the combination of the 5'-end sequence/3'-end sequence is selected from those defined in the specification. The primer sets can be used in antisense therapy and in gene therapy. The primers are useful for synthesizing polynucleotides, particularly full-length cDNAs. The primers are also useful for the detection and/or diagnosis of the abnormality of the proteins encoded by the full-length cDNAs. The primers allow obtaining of the full-length cDNAs easily without any specialised methods. AAH03166 to AAH13628 and AAH13633 to AAH18742 represent human cDNA sequences; AAH92446 to AAH95893 represent human amino acid sequences; and AAH13629 to AAH13632 represent oligonucleotides, all of which are used in the exemplification of the present invention

XX Sequence 2993 BP; 666 A; 864 C; 856 G; 607 T; 0 U; 0 Other;

## Alignment Scores:

Pred. No.: 1.33e-282 Length: 2993  
 Score: 295.00 Matches: 681  
 Percent Similarity: 98.84% Conservatives: 0  
 Best Local Similarity: 98.84% Mismatches: 4  
 Query Match: 40.08% Indels: 8  
 DB: 4 Gaps: 0

US-10-045-815-4 (1-736) x AAH16288 (1-2993)

Qy 1 MetAlaValArgAlaLeuLysLeuThrThrLeuLeuAlaValAlaAlaSer 20  
 Db 42 ATGGCGGTACCGCGGTGAACCTGCTGACCACTGCTGGCTGCGGGCGGCTGCTCC 101  
 Qy 21 GlnAlaGluValGluSerGluAlaGlyTrpGlyMetValThrProAspLeuPheAla 40  
 Db 102 CAAGCCGAGGTGCGAGTCCGAGGAGGATGGGCATGCTGAGCGCTGATCTGCTTCGCC 161  
 Qy 41 GluGlyThrAlaAlaValArgAlaGlyAspTrpProGlyValValLeuSerMetGluArg 60  
 Db 162 GAGGGGACCGGAGCTTACGCGCGGGAGCTGGCCGGGGTGGTCTTGAGCATGGAAACGG 251  
 Qy 61 AlaLeuArgSerArgAlaAlaLeuArgAlaLeuArgLeuArgCysArgThrGlnCysAla 80  
 Db 222 GCGCTGCGCTCCCGGCGAGCCCTCCGCGCTTCCGCTGCGCTGCCGCACCCAGTGTGCC 281  
 Qy 81 AlaAspPheProTrpGluLeuAspProAspTrpSerProSerProAlaGlnAlaSerGly 100  
 Db 282 GCCGACTTCCGCTGGGAGCTGGACCCGACTGGTTCGCCCGCGCGGCCCGGCTCGGGC 341  
 Qy 101 AlaGly-AlaLeuArgAspLeuSerPhePheGlyGlyLeuLeuArgArgAlaAlaCysLe 120  
 Db 342 GCCGC-CGCCCTGGCGACCTGAGCTTCTTCGGGGGCTTCTGCGTGGCTGCGCTGCCCT 400  
 Qy 120 uArgArgCysLeuGlyProProAlaAlaHisSerLeuSerGluLeuMetGluLeuGluPh 140  
 Db 401 GCGCCGCTGCTCCGGCGCGCGCGCCCTCGCTCAGCGAAAGAGATGGAGTGGAGTT 460  
 Qy 140 eArgLysArgSerProTrpAsnTyrLeuGlnValAlaTyrPheLysIleAsnLysLeuGlu 160  
 Db 461 CCGCAAGCGAGCCCTTACAACTACCTGAGGTGCGCTTCTCAAGATCAACAAAGTTGGA 520  
 Qy 160 uLysAlaValAlaAlaAlaHisThrPhePheValGlyAsnProGluHisMetGluMetGlu 180  
 Db 521 GAAAGCTGTGCTGCGAGCACACACTTCTTCGTGGGCAATCTCTGAGCACATGGAATGCA 580  
 Qy 180 nGlnAsnLeuAspTyrTrpGlnThrMetSerGlyValLysGluAlaAspPheLysAspLe 200  
 Db 581 GCAGAACTAGACTATTACCAACCACTGCTGGAGTGAAGAGGCCGACTTCAAGGATCT 640  
 Qy 200 uGluThrGlnProHisMetGlnGluPheArgLeuGlyValArgLeuTyrSerGluGlu 220  
 Db 641 TGAGACTCAACCCCATATGCAAGAAATTCAGACTGGGAGTCCGACTCTTACTCAGAGGAACA 700

Qy 220 nProGlnGluAlaValProHisLeuGluAlaAlaLeuGlnGluTyrPheValAlaTyrGlu 240  
 Db 701 GCCACAGGAGCTGTGCCCCACCTAGAGCGGGCTGCAAGATACCTTTGTGGCCATGA 760  
 Qy 240 uGluCysArgAlaLeuCysGluGlyProTyrAspTyrAspGlyTyrAsnTyrLeuGluTyr 260  
 Db 761 GGAGTGGCGTGGCTTCTGCGAAGGCCCTATGACTACGATGGCTACCACTACCTTGAGTA 820  
 Qy 260 rAsnAlaAspLeuPheGlnAlaIleThrAspHisTyrIleGlnValLeuAsnCysLysGlu 280  
 Db 821 CAACGCTGACCTTCTCCAGGCCATCACAGATCATTCATCCAGGTCTCACTCACTGAACA 880  
 Qy 280 nAsnCysValThrGluLeuAlaSerHisProSerArgGluLysProPheGluAspPheLe 300  
 Db 881 GAACTGTGTACGAGCTTGTCTCCACCCCAAGTCAGAGAAGCCCTTTGAAGACTTCTCT 940  
 Qy 300 uProSerHisTyrAsnTyrLeuGlnPheAlaTyrTyrAsnIleGlyAsnTyrThrGlnAl 320  
 Db 941 CCCATCGCATTAATAATTATCTGCAGTTTGCCTTATTAACATTTGGAAATTATCACAGGC 1000  
 Qy 320 agly-GluCysAlaLysThrTyrLeuLeuPhePheProAsnAspGluValMetAsnGlnA 340  
 Db 1001 TGT-TGAATGTGCCAAGACCTATCTTCTTCTTCCCAATGACGAGGTGATGAACAAA 1059  
 Qy 340 snLeuAlaTyrTyrAlaAlaMetLeuGlyGluGluHisThrArgSerIleGlyProArg 360  
 Db 1060 ATTTGGCTTATTATGAGCTATGCTTGGAGAGAACACACAGATCCATCGGCCCGCTG 1119  
 Qy 360 LuSerAlaLysGluTyrArgGlnArgSerLeuLeuGluLysGluLeuLeuPheAlaT 380  
 Db 1120 AGAGTGCACAGAGTACCGACAGGACCTACTTGGAAAAGAACTGCTTTCTTCGCTT 1179  
 Qy 380 yAspValPheGlyIleProPheValAspProAspSerTyrThrProGlu-GluValIle 399  
 Db 1180 ATGATGTTTTTGGAAATCTTGTGGATCCGGATTCATGGACTCCAGG-AGAAGTGA 1238  
 Qy 400 ProLysArgLeuGlnGluLysGlnLysSerGluArgGluThrAlaValArgIleSerGln 419  
 Db 1239 CCCAAGAGATTGCAAGAGAAACAGAGTCAGACGGGAAACAGCCGTACGATCTCCAG 1298  
 Qy 420 GluIleGlyAsnLeuMetLysGluIleGluThrLeuValGluGluLysThrLysGluSer 439  
 Db 1299 GAGATTGGGAACCTTATGAAGAAATCAGACCTTGTGGAAAGAGAGACCAAGAGTCA 1358  
 Qy 440 LeuAspValSerArgLeuThrArgGluGlyGlyProLeuLeuTyrGluGlyLysSerLeu 459  
 Db 1359 CTGGATGTGAGCAGACTACCCGGAGGTGGCCCTCTGCTGTATGAAGCATCAGTCTC 1418  
 Qy 460 ThrMetAsnSerLysLeuLeuAsnGlyTyr-GlnArgValValMetAspGlyValIleSe 479  
 Db 1419 ACCATGAATCCAAACTCCTGAATGGTTC-CCAGCGGTGGTGTGGACGGCTAATCTC 1477  
 Qy 479 rAspHisGluCysGlnGluLeuGlnArgLeuThrAsnValAlaAlaThrSerGlyAspGlu 499  
 Db 1478 TGACCACAGTGTACAGGAGCTGACAGACTGACCAATGTGGCAGCAACCTCAGAGATGG 1537  
 Qy 499 yTyrArgGlyGlnThrSerProHisThrProAsnGluLysPheTyrGlyValThrValPh 519  
 Db 1538 CTACCGGGTCCAGACTCCCCACATCTCCCAATGAAAAGTTCTATGGTGTCTGCTCTT 1597  
 Qy 519 elysAlaLeuLysLeuGlyGlnGluGlyLysValProLeuGlnSerAlaHisLeuTyrTy 539  
 Db 1598 CAAAGCCCTCAAGCTGGGGCAAGGCAAGTTCCTCTGCAGAGTCCCACTGCTGTACTA 1657  
 Qy 539 rAsnValThrGluLysValArgArgIleMetGluSerTyrPheArgLeuAspThrProLe 559  
 Db 1658 CAACGTGACGGAAGAGTGGCGGCATCATCGAGTCTCTACTTCCGCCCTGGATACGCCCT 1717  
 Qy 559 uTyrPheSerTyrSerHisLeuValCysArgThrAlaIleGluGluValGlnAlaGluAr 579  
 Db 1718 CTACTTTTCTACTCTCATCTGCTGTCGCACTGCCATCGAAGAGGTCCAGGAGAGAG 1777

QY 579 gLyAspAspSerHisProValHisValAspAsnCysIleLeuAsnAlaGluThrLeuVa 599  
Db 1778 GAAGGATGATGTCATCCAGTCCACGTGGACCACTGCATCCTGAAATGCCGAGACCTCGT 1837  
QY 599 lCysValLysGluProProAlaTyrThrPheArgAspTyrSerAlaIleLeuTyrLeuAs 619  
Db 1838 GTGTGTCAAGAGCCGCCAGCTACACCTTCGCGACTACAGCGCCATCCTTACCTAAA 1897  
QY 619 nGlyAspPheAspGlyGlyAsnPheTyrPheThrGluLeuAspAlaLysThrValThrAl 639  
Db 1898 TGGGGACCTTCGATGGCGGAACCTTTATTTCACTGAACCTGGATGCCCAACACCGTGACGCG 1957  
QY 639 aGluValGlnProGlnCysGlyArgAlaValAlaGlyPheSerSerGlyThrGluAsnProHi 659  
Db 1958 AGAGGTGCACCTCAGTGTGGAGAGCCGTGGGATTTCTCTCAGGCACCTGAAACCCACA 2017  
QY 659 sGlyValLysAlaValThrArgGlyGlnArgCysAlaIleAlaLeuTyrPheThrLeuAs 679  
Db 2018 TGGAGTGAAGCGTGTCCACAGGGCGGCGCTGTGCATCGCCCTGTGGTTTCACCCCTGA 2077  
QY 679 pProArgHisSerGluArg 685  
Db 2078 CCCTCGACACAGCGAGCGG 2096

## RESULT 13

AA517572  
ID AA517572 standard; cDNA; 2127 BP.

AC AA517572;

DT 26-FEB-2002 (first entry)

XX DNA encoding novel secreted protein #1.

XX Secreted protein; cytostatic; immunosuppressive; vulnery; vaccine;  
KW antiinflammatory; neuroprotective; nephrotropic; cardiovascular; human;  
KW cancer; autoimmune disease; wound healing disorder; infection;  
KW haematopoietic disorder; inflammatory disorder; infertility;  
KW neurological disease; psychiatric disease; cardiovascular disease;  
KW respiratory disease; renal; gastrointestinal; ss.  
XX Homo sapiens.

XX Key Location/Qualifiers

FF CDS 1..2127

FT FT /\*tag= a

FT FT /product= "Human secreted protein"

XX WO200179454-A1.

XX 25-OCT-2001.

XX 11-APR-2001; 2001WO-US011797.

XX 13-APR-2000; 2000US-0196603P.

XX 24-APR-2000; 2000US-0199417P.

XX (SMIK ) SMITHKLINE BEECHAM CORP.

XX (SMIK ) SMITHKLINE BEECHAM PLC.

XX Agarwal P, Murdoch PR, Rizvi SK, Smith RF, Xiang Z;

XX WPI; 2002-061975/08.

XX P-PSDB; AAU09860.

XX New secreted proteins or polypeptides, useful for treating e.g. cancer,  
PT autoimmune diseases, wound healing disorder, infections, hematopoietic  
PT disorders, inflammatory disorders, infertility, cancer.

XX Claim 2; Page 34-35; 92pp; English.

XX The invention relates to an isolated novel secreted polypeptide (I) and  
CC polynucleotide (II). (I) and (II) are useful for treating cancer,

CC autoimmune diseases, wound healing disorder, infections, haematopoietic  
CC disorders, inflammatory disorders, infertility, neurological and  
CC psychiatric diseases, cardiovascular diseases, respiratory diseases,  
CC renal diseases, or gastrointestinal diseases. These may also be used to  
CC treat diseases, abnormalities and disorders caused by abnormal  
CC expression, production, function and/or metabolism of the genes, as  
CC vaccines for inducing immunological response in a mammal, and in  
CC screening methods for detecting the effect of added compounds on the  
CC production of mRNA and polypeptide in cells. The polypeptides can be used  
CC as immunogens to produce antibodies immunospecific for the polypeptides,  
CC and to identify membrane-bound or soluble receptors. The polynucleotides,  
CC may be used as diagnostic reagents, in chromosome localisation studies,  
CC and in tissue expression studies. The present sequence represents the  
CC coding sequence of novel human secreted protein #1

XX SQ Sequence 2127 BP; 495 A; 623 C; 589 G; 420 T; 0 U; 0 Other;

## Alignment Scores:

Pred. No.: 7.31e-263 Length: 2127  
Score: 275.00 Matches: 375  
Percent Similarity: 99.47% Conservative: 0  
Best Local Similarity: 99.47% Mismatches: 1  
Query Match: 27.36% Indels: 2  
DB: 6 Gaps: 0

US-10-045-815-4 (1-736) x AA517572 (1-2127)

QY 361 SerAlaLysGluTyrArgGlnArgSerLeuLeuGluLysGluLeuPhePheAlaTyr 380  
Db 997 AGTGCCAGAGGAGTACCGACAGCGAGCTACTGGAAAAGAACTCTTTCTTCGTTAT 1056  
QY 381 AspValPheGlyIleProPheValAspProAspSerTrpThrProGluGluValIlePro 400  
Db 1057 CATGTTTTTGGAAATTCCTTTTGTGATCCGGATTTCATGCACTCCAGAGAAAGTATTC 1116  
QY 401 LysArgLeuGlnGluLysGlnLysSerGluArgGluThrAlaValArgIleSerGlnGlu 420  
Db 1117 AAGAGATTGCAAGAGAAACAGAGTACAGACGGAAACAGCGGTACCGATTCTCCAGAG 1176  
QY 421 IleGlyAsnLeuMetLysGluIleGluThrLeuValGluGluLysThrLysGluSerLeu 440  
Db 1177 ATTGGGAACCTTATGAAGGAAATCGAGACCTTTGTGGAAAGAGAGACCAAGAGTCACTG 1236  
QY 441 AspValSerArgLeuThrArgGluGlyGlyProLeuLeuTyrGluGlyIleSerLeuThr 460  
Db 1237 GATGTGACGACACTGACCCGGAAGGTGCCCCCTGCTGTATGAAGGCATCAGTCTCACC 1296  
QY 461 MetAsnSerLysLeuLeuAsnGlyTyr-GlnArgValValMetAspGlyValIleSerAs 480  
Db 1297 ATGAATCTCAAACTCTTGAATGGTTC-CCAGCGGGTGTGTATGGACGGGTAACTCTGA 1355  
QY 480 pHisGluCysGlnGluLeuGlnArgLeuThrAsnValAlaAlaThrSerGlyAspGlyTy 500  
Db 1356 CCACGAGTGTCCAGAGCTGCAGAGACTGCCAATGTGGCAGCAACCTCAGGAGATGGCTA 1415  
QY 500 rArgGlyGlnThrSerProHisThrProAsnGluLysPheTyrGlyValThrValPheIy 520  
Db 1416 CCGGGGTGAGACCTCCCACTATCCCAATGAAGAATTCATGGTGTCTGCTGCTTCAA 1475  
QY 520 sAlaLeuLysLeuGlyGlnGluLysValProLeuGlnSerAlaHisLeuTyrTyrAs 540  
Db 1476 AGCCCTCAAGCTGGGGCAAGAGGCAAAAGTTCTCTGCGAGAGTGGCCACCTGATACAA 1535  
QY 540 nValThrGluLysValArgArgIleMetGluSerTyrPheArgLeuAspThrProLeuTy 560  
Db 1536 CGTGACGAGAGAGGTGCGGCGCATCATGGAGTCTTCTTCCGCTGGATACGCCCTCTA 1595  
QY 560 rPheSerTyrSerHisLeuValCysArgThrAlaIleGluGluValGlnAlaGluArgLy 580  
Db 1596 CTTTTCTCTACTCTCATCTGGTGTGCGCCTGCCATCGAAGAGGTCCAGGCAGAGGAA 1655  
QY 580 sAspAspSerHisProValHisValAspAsnCysIleLeuAsnAlaGluThrLeuValCy 600

1656 CGATGATAGTATCCAGTCCAGCTGAGCACTGATCTGATGCGGAGACCTCGTGTG 1715  
QY 600 sValylsGluProProlaTyThrPheArgAspTy-SerAlaIleLeuTyRleuAsnGI 620  
DB 1716 TGTCAAGAGAGCCCGAGCTACACTTCGCGAGCTACAGCGCCATCTTTACCTAAATGG 1775  
QY 620 YAspPheAspGlyGlyAsnPheTyThrGluLeuAspAlaLysThrValThrAlaGI 640  
DB 1776 GGACTTCGATCGCGAAACTTTTATTACCTGAATGATGCCAAGACCGTGAGCGGAGA 1835  
QY 640 uValGlnProGlnCysGlyArgAlaValGlyPheSerSerGlyThrGluAsnProHisGI 660  
DB 1836 GGTGAGCCCTCAGTGTGGAAGAGCGGTGGATCTCTTCAGGCACTGAAACCCACATGG 1895  
QY 660 YValylsAlaValThrArgGlyGlnArgCysAlaIleAlaLeuTrpPheThrIleuAspPr 680  
DB 1896 AGTGAAGCTGTCCACAGGGGCGAGCGCTGTGCCATCGCCCTGTGGTTACCTGGACCC 1955  
QY 680 oArgHisSerGluArgAspArgValGlnAlaAspLeuValLysMetLeuPheSerPr 700  
DB 1956 TCGACAGCGAGCGGAGACGGGTGCAGGACAGATGACTGTGTGAAGATGCTTTCAGCCC 2015  
QY 700 oGluGluMetAspLeuSerGlnGluGlnProLeuAspAlaGlnGlnGlyProProGluPr 720  
DB 2016 AGAAGAGATGGACCTCTCCAGGAGAGACCCCTGGATGCCCGAGCGGTCCCGCCGAACC 2075  
QY 720 oAlaGlnGluSerLeuSerGlySerGluSerLysProLysAspGluLeu 736  
DB 2076 TGCACAGAGTCTCTTCAGGACAGTGAATCGAAGCCCGAGGATGAGCTA 2124

RESULT 14  
ABZ11345  
ID ABZ11345 standard; cDNA; 2152 BP.  
XX AC ABZ11345;  
XX 20-JAN-2003 (first entry)  
XX Human polynucleotide SEQ ID NO 227.  
XX Human; genome mapping; gene therapy; food supplement; virus; fungus;  
XX cell-proliferative disorder; neurodegenerative disease; bacterial;  
XX Parkinson's disease; Alzheimer's disease; autoimmune disease;  
XX multiple sclerosis; diabetes; genetic disorder; wound; burn; infection;  
XX arthritis; cyostatic; immunomodulator; nootropic; neuroprotective;  
XX antiparkinsonian; antidiabetic; immunosuppressive; dermatological;  
XX haemostatic; vulnery; fungicide; antibacterial; virucide; protozoacide;  
XX antiarthritic; gene; ss.  
XX Homo sapiens.  
OS  
XX WO200270539-A2.  
FN  
XX 12-SEP-2002.  
PD  
XX 05-MAR-2002; 2002WO-US005095.  
PF  
XX 05-MAR-2001; 2001US-00799451.  
PR  
XX (HYSE-) HYSEQ INC.  
PA  
XX Tang YT, Zhou P, Goodrich RW, Asundi V, Zhang J, Zhao OA, Ren F;  
PI Xue AJ, Yang Y, Ma Y, Yamazaki V, Chen R, Wang Z, Ghosh M;  
PI Wehrman T, Wang J, Wang D, Drmanac RT;  
XX WPI; 2002-759812/82.  
DR P-ESDB; ABP69128.  
DR  
XX New polynucleotides comprising sequences assembled from expressed  
PT sequence tags (ESTs), useful for treating cell-proliferative,  
PT neurodegenerative, autoimmune, genetic, myeloid or lymphoid, or platelet  
PT or coagulation disorders.  
XX

PS Claim 1; SEQ ID NO 227; 1012bp + Sequence Listing; English.  
XX  
The invention relates to an isolated polynucleotide (I) comprising a  
nucleotide sequence selected from any of 948 sequences (ABZ11119-  
ABZ12066) or their mature protein coding portion, active domain coding  
protein or complementary sequences. The polynucleotides are useful for  
identifying expressed genes or for physical mapping of human genome. The  
encoded polypeptides (ABP6902-ABP6949) are useful as molecular weight  
markers, as a food supplement, for generating antibodies, in medical  
imaging, screening and diagnostic assays and for treating cell-  
proliferative disorders (cancer), neurodegenerative diseases (Parkinson's  
or Alzheimer's disease), autoimmune diseases (multiple sclerosis,  
diabetes, lupus) genetic disorders, myeloid or lymphoid disorders,  
platelet or coagulation disorders, wound, burns, incision, ulcers, liver  
or lung fibrosis, infections (bacterial, viral, fungal, parasitic),  
arthritis, etc. Note: The sequence data for this patent did not form part  
of the printed specification, but was obtained in electronic format  
directly from WIPO at ftp.wipo.int/pub/published\_pct\_sequences  
XX  
SQ Sequence 2152 BP; 494 A; 636 C; 602 G; 420 T; 0 U; 0 Other;  
Alignment Scores:  
Pred. No.: 4.8e-232 Length: 2152  
Score: 244.00 Matches: 358  
Percent Similarity: 98.90% Conservative: 0  
Best Local Similarity: 98.90% Mismatches: 2  
Query Match: 33.15% Indels: 4  
DB: 6 Gaps: 0  
US-10-045-815-4 (1-736) x ABZ11345 (1-2152)  
QY 1 MetAlaValArgAlaLeuLysLeuThrThrLeuLeuAlaValAlaAlaSer 20  
DB 32 ATGGCGGTACGGCGCTTGAAGCTGCTGACACACACTGCTGGCTGCTGGCGCTGCCTCC 91  
QY 21 GlnAlaGluValGluSerGluAlaGlyTrpGlyMetValThrProAspLeuPheAla 40  
DB 92 CAAGCGAGGTGAGTCCGAGGAGGATGGGGCATGGTGACGCTGATCTCTCTTCGCC 151  
QY 41 GluGlyThrAlaAlaTyzAlaArgGlyAspTrpProGlyValValLeuSerMetGluArg 60  
DB 152 GAGGGAGCGGAGCTACGCGCGGGGAGCTGGCCGGGGTGGTCTCGAGCATGGACGG 211  
QY 61 AlaLeuArgSerArgAlaAlaLeuArgAlaLeuArgLeuArgCysArgThrGlnCysAla 80  
DB 212 GCGCTGCGCTCCCGGAGCGCTCCGCGCCCTTCCGCTGCGCTGCGCACCACCGCTGTC 271  
QY 81 AlaAspPheProTrpGluLeuAspProAspTrpSerProAlaGlnAlaSerGly 100  
DB 272 GCGGACTTCCGTTGGGAGCTGGACCCCGACTGGTCCCCAGCCGCGCCAGCTCGGGC 331  
QY 101 AlaGly-AlaLeuArgAspLeuSerPhePheGlyGlyLeuLeuArgArgAlaAlaCysLe 120  
DB 332 GCGGC-CGCGCTGCGGAGCTGAGCTTCTTGGGGGCGCTTCTGGGCTCGCGCTGCGCT 390  
QY 120 uArgArgCysLeuGlyProProAlaAlaHisSerLeuSerGluGluMetGluLeuGluPh 140  
DB 391 GCGCGCTGCTTCGGGCGCGCGCGCCACTCGCTCAGCGAAGAGATGGAGCTTGGAGTT 450  
QY 140 eArgLysArgSerProTyzAsnTyzLeuGlnValAlaTyzPheLysLeuLeuGlu 160  
DB 451 CGCAAGCGGAGCGCTCAACTACCTACCTGCGCTTCTTCAAGATCAACAAGTTGGA 510  
QY 160 uLysAlaValAlaAlaAlaHisThrPhePheValGlyAsnProGluHisMetGluMetGI 180  
DB 511 GAAAGCTGTGTGCGAGCACACACTTCTTGTGGGCAATCTTGAGCACATGGAATGCA 570  
QY 180 nGlnAsnLeuAspTyzTyzGlnThrMetSerGlyValLysGluAlaAspPheLysAspLe 200  
DB 571 GCAGAACCTAGACTATTACCAAAACCATGCTCGAGTGAAGGAGGCGGACTTCAAGGATCT 630  
QY 200 uGluThrGlnProHisMetGlnGluPheArgLeuGlyValArgLeuTyzSerGluGluGI 220

Db 631 TGAGACTCAACCCATATGCAAGAAATTTTCGACTGGGAGTGCAGCTCTACTCAGAGGAACA 690  
Qy 220 nProGlnGluAlaValProHisLeuGluAlaLeuGlnGluTyrPheValAlaTyrG1 240  
Db 691 GCCACAGGAACCTGTGCCCCACCTAGAGCGCGCTGCAGAGATATCTTTGTGGCCCTATGA 750  
Qy 240 uGluCysArgAlaLeuCysGluGlyProTyrAspTyrAspGlyTyrAsnTyrLeuGluTyr 260  
Db 751 GGAGTGCCTGCTCTGCGAAGGCGCTATGACTACGATGGCTACAACTACCTTGAAGTA 810  
Qy 260 rAsnAlaAspLeuPheGlnAlaIleThrAspHisTyrIleGlnValLeuAsnCysIysG1 280  
Db 811 CAACGCTGACCTTTCAGAGCCATCAGATCATTATCATCCAGTCTCAACTGTAAAGCA 870  
Qy 280 nAsnCysValThrGluLeuAlaSerHisProSerArgGluLysProPheGluAspPheLe 300  
Db 871 GAACGTGTACAGGAGCTTCTCCACCCCAAGTCGAGAGAACCCCTTTGAAGACTTCCT 930  
Qy 300 uProSerHisTyrAsnTyrLeuGlnPheAlaTyrTyrAsnIleGlyAsnTyrThrGlnAl 320  
Db 931 CCCATCGCATATATATCTGAGCTTTCCTACTATTAACATTTGGGAATATATACACAGGC 990  
Qy 320 aGly-GluCysAlaLysThrTyrLeuLeuPhePheProAsnAspGluValMetAsnGlnA 340  
Db 991 TGT-TGATGTGCAGAGACCTATCTTCTTCTTCCCAATGACGAGGTGATGACCAAA 1049  
Qy 340 snLeuAlaTyrTyrAlaAlaMetLeuGlyGluGluHisThrArgSerIleGlyProArgG 360  
Db 1050 ATATTGGCTATTATGACGCTATGCTGGAGAGAGACACACAGATCATCTCGGCCCGCTG 1109  
Qy 360 lu 360  
Db 1110 AG 1111

RESULT 15  
ABZ11344  
ID ABZ11344 standard; cDNA; 1309 BP.  
XX AC ABZ11344;  
XX DT 20-JAN-2003 (first entry)  
XX DE Human polynucleotide SEQ ID NO 226.  
XX EX  
XX KW Human; genome mapping; gene therapy; food supplement; virus; fungus;  
XX KW cell-proliferative disorder; neurodegenerative disease; bacterial;  
XX KW Parkinson's disease; Alzheimer's disease; autoimmune disease;  
XX KW multiple sclerosis; diabetes; genetic disorder; wound; burn; infection;  
XX KW arthritis; cytostatic; immunomodulator; nootropic; neuroprotective;  
XX KW antiparkinsonian; antidiabetic; immunosuppressive; dermatological;  
XX KW haemostatic; vulnery; fungicide; antibacterial; virucide; protozoacide;  
XX KW antiarthritic; Gene; ss.  
XX OS Homo sapiens.  
XX PN WO200270539-A2.  
XX PD 12-SEP-2002.  
XX PF 05-MAR-2002; 2002WO-US0005095.  
XX PR 05-MAR-2001; 2001US-00799451.  
XX PA (HYSE-) HYSEQ INC.  
XX PI Tang YT, Zhou P, Goodrich RW, Asundi V, Zhang J, Zhao QA, Ren F;  
XX PI Xue AU, Yang Y, Yamazaki V, Chen R, Wang Z, Ghosh M;  
XX PI Wehrman T, Wang J, Wang D, Drmanac RT;  
XX DR WPI; 2002-759812/82.  
XX DR P-PSDB; ABP69127.  
XX PT New polynucleotides comprising sequences assembled from expressed

PT sequence tags (ESTs), useful for treating cell-proliferative,  
PT neurodegenerative, autoimmune, genetic, myeloid or lymphoid, or platelet  
XX or coagulation disorders.  
PS Claim 1; SEQ ID NO 226; 1012pp + Sequence Listing; English.  
XX  
CC The invention relates to an isolated polynucleotide (I) comprising a  
CC nucleotide sequence selected from any of 948 sequences (ABZ11119-  
CC ABZ12066) or their mature protein coding portion, active domain coding  
CC protein or complementary sequences. The polynucleotides are useful for  
CC identifying expressed genes or for physical mapping of human genome. The  
CC encoded polypeptides (ABP68902-ABP69849) are useful as molecular weight  
CC markers, as a food supplement, for generating antibodies, in medical  
CC imaging, screening and diagnostic assays and for treating cell-  
CC proliferative disorders (cancer), neurodegenerative diseases (Parkinson's  
CC or Alzheimer's disease), autoimmune diseases (multiple sclerosis,  
CC diabetes, lupus) genetic disorders, myeloid or lymphoid disorders,  
CC platelet or coagulation disorders, wound, burn, incision, ulcers, liver  
CC or lung fibrosis, infections (bacterial, viral, fungal, parasitic),  
CC arthritis, etc. Note: The sequence data for this patent did not form part  
CC of the printed specification, but was obtained in electronic format  
CC directly from WIPO at ftp.wipo.int/pub/published\_pct\_sequences  
XX  
SQ Sequence 1309 BP; 287 A; 384 C; 365 G; 273 T; 0 U; 0 Other;  
Alignment Scores:  
Pred. No.: 2.12e-207 Length: 1309  
Score: 219.00 Matches: 406  
Percent Similarity: 99.02% Conservative: 0  
Best Local Similarity: 99.02% Mismatches: 2  
Query Match: 29.76% Indels: 4  
DB: 6 Gaps: 0  
US-10-045-815-4 (1-736) x ABZ11344 (1-1309)  
Qy 1 MetAlaValArgAlaLeuLysLeuLeuThrThrLeuLeuAlaValAlaAlaSer 20  
Db 32 ATGGCGGTACGGCGTTCGAGCTGCTGACCACTGCTGGCTGCTGCGCTGCCCTCC 91  
Qy 21 GlnAlaGluValGluSerGluAlaGlyTrpGlyMetValThrProAspLeuPheAla 40  
Db 92 CAAGCCGAGGTGAGTCCGAGCGAGGATGGGGCATGGTGACGCCCTGATCTCTCTCC 151  
Qy 41 GluGlyThrAlaAlaTyrAlaArgGlyAspTrpProGlyValValLeuSerMetGluArg 60  
Db 152 GAGGGGACCGAGCCCTACGCGCGGGGAGCTGGCCGGGGTGGTCTCTGAGGATGGACGG 211  
Qy 61 AlaLeuArgSerArgAlaAlaLeuArgAlaLeuArgLeuArgCysArgThrGlnCysAla 80  
Db 212 GCGCTGCGTCCGCGGACGCTCCGCGCCCTTCGCTGCGCTGCGCACCCAGGTGCC 271  
Qy 81 AlaAspPheProTrpGluLeuAspProAspTrpSerProSerProAlaGlnAlaSerGly 100  
Db 272 GCCGACTTCCCGTGGGAGCTGAGCCCGACCTGGTCCCGCCAGCCCGCCAGGCTTCGGGC 331  
Qy 101 AlaGly-AlaLeuArgAspLeuSerPhePheGlyGlyLeuLeuArgArgAlaAlaCysLe 120  
Db 332 GCGGC- GCGCCCTGCGCGACCTGAGCTTCTTCGGGGGCTCTTCGCTGCGCTGCGCT 390  
Qy 120 uArgArgCysLeuGlyProProAlaAlaHisSerLeuSerGluGluMetGluLeuGluPh 140  
Db 391 GCGCGCTGCTCCGCGCGCGCCGCGCCCTACGCTCAGCGAGAGATGGAGCTGAGTT 450  
Qy 140 eArgLysArgSerProTyrAsnTyrLeuGlnValAlaTyrPheLysIleAsnLysLeuG1 160  
Db 451 CGCGAAGCGGAGCCCTTACACTACCTCAGGTGCGCTACTTCAAGATCAACAAGTTGGA 510  
Qy 160 uLysAlaValAlaAlaAlaHisThrPhePheValGlyAsnProGluHisMetGluMetG1 180  
Db 511 GAAAGCTGTGCTGCAGCACACACCTTCTTCGTGGGCAATCCTTGAGCACATGGAAATGCA 570  
Qy 180 nGlnAsnLeuAspTyrTyrGlnThrMetSerGlyValLysGluAlaAspPheLysAspLe 200

Db 571 GCAGAACCTAGACTATTACCAACCATGCTCTGGAGTGAAGAGGCCGACTTCAAGGATCT 630  
Qy 200 uGluThrGlnProHisMetGlnGluPheArgLeuGlyValArgLeuTyrSerGluGluG1 220  
Db 631 TGAGACTCAACCCCATATGCAAGAAATTCGACTGGGAGTGGACTCTACTCAGAGGAACA 690  
Qy 220 nProGlnGluAlaValProHisLeuGluAlaAlaLeuGlnGluTyrPheValAlaTyrG1 240  
Db 691 GCCACAGGAAGCTGTGCCACCTAGAGGGGGCGCTCAAGAATACCTTTGTGGCCCTATGA 750  
Qy 240 uGluCysArgAlaLeuCysGluGlyProTyrAspTyrAspGlyTyrAsnTyrLeuGluTyr 260  
Db 751 GGAGTGGCGTGCCTCTGGAGAGGGCCCTATGACTACGATGGCTACAACTACCTTGAGTA 810  
Qy 260 rAsnAlaAspLeuPheGlnAlaIleThrAspHisTyrIleGlnValLeuAsnCysLysG1 280  
Db 811 CAAGCTGACCTCTTCCAGGCCATCACAGATCATTACATCCAGTCCCTCAACTGTAAGCA 870  
Qy 280 nAsnCysValThrGlnLeuAlaSerHisProSerArgGluLysProPheGluAspPheLe 300  
Db 871 GAAGTGTACGGAGCTTGCTTCCACCCCAAGTCGAGAGAGCCCTTTGAAGACTTCCT 930  
Qy 300 uProSerHisTyrAsnTyrLeuGlnPheAlaTyrTyrAsnIleGlyAsnTyrThrGlnAl 320  
Db 931 CCCATCGCATTATTAATTATCTGCAGTTTGGCTACTATAACATTGGGAATTATACACAGGC 990  
Qy 320 aGly-GluCysAlaLysThrTyrLeuLeuPhePheProAsnAspGluValMetAsnGlnA 340  
Db 991 TGT-TGAATGTCACAGACCTATCTTCTCTTCCCAATGACGAGGTGATGAACCAA 1049  
Qy 340 snLeuAlaTyrTyrAlaAlaMetLeuGlyGluGluHisThrArgSerIleGlyProArgG 360  
Db 1050 ATTTGGCTATTATGCAGCTATGCTTGAGAGAAGACACACAGATCCATCGGCCCCCGTG 1109  
Qy 360 luSerAlaLysGluTyrArgGlnArgSerLeuLeuGluLysGluLeuLeuPhePheAlaT 380  
Db 1110 AGAGTGCCAAAGGAGTACCGACAGCGAGCCTACTGGAAAAAGAACTGCTTTCTTCGCTT 1169  
Qy 380 yrAspValPheGlyIleProPheValAspProAspSerTyrThrProGluGluValIleP 400  
Db 1170 ATGATGTTTTGGAAATTCCTTTGTGGATCGGATTCATGGACTCCAGAGAAGAGTGATTC 1229  
Qy 400 roLysArgLeuGlnGluLysGlnLys 408  
Db 1230 CCAAGAGATTGCAAGAGAAACAGAAG 1255

Search completed: July 18, 2004, 18:38:50  
Job time : 1051 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - nucleic search, using frame\_plus\_p2n model

Run on: July 18, 2004, 18:22:31 ; Search time 1081 Seconds  
(without alignments)  
3321.616 Million cell updates/sec

Title: us-10-045-815-4  
Perfect score: 736  
Sequence: 1 MAVRALKLLTLLAVVAAS.....PPEPAQESLSGSEKPKDEL 736

Scoring table:

OLIGO  
Xgapop 60.0 , Xgapext 60.0  
Ygapop 60.0 , Ygapext 60.0  
Fgapop 6.0 , Fgapext 7.0  
Delop 6.0 , Delext 7.0

Searched: 3190392 seqs, 2439311697 residues

Word size: 1

Total number of hits satisfying chosen parameters: 6373542

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Command line parameters:

-MODEL=frame+ p2n.model -DEV=xlp  
-Q=/cgn2\_1/USPTO\_spool\_p/US10045815/runat\_14072004\_123120\_17294/app\_query.fasta\_1.903  
-DB=Published Applications NA -QFWT=fastap -SURFIX=oli.rnpb -MINMATCH=0.1  
-LOOFCU=0 -LOOPEXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=oligo  
-TRANS=human40.cdi -LIST=45 -DOCALIGN=200 -THR\_SCORE=quality -THR\_MIN=1  
-ALIGN=15 -MODE=LOCAL -OUTFWT=pto -NORM=ext -HRAPISE=500 -MINLEN=0  
-MAXLEN=2000000000 -USER=US10045815 @CGN 1 1 912 @runat\_14072004\_123120\_17294  
-NCPU=6 -ICPU=3 -NO\_MMAP -LARGEQUERY -NEG\_SCORES=0 -WAIT -DSPLOCK=100  
-LONGLOG -DEV\_TIMEOUT=120 -WARN\_TIMEOUT=30 -THREADS=1 -XGAPOP=60 -XGAPEXT=60  
-FGAPOP=6 -FGAPEXT=7 -YGAPOP=60 -YGAPEXT=60 -DELOP=6 -DELEXT=7

Database : Published Applications NA:  
1: /cgn2\_6/ptodata/2/pubpna/US07\_PUBCOMB.seq:  
2: /cgn2\_6/ptodata/2/pubpna/PCT\_NEW\_PUB.seq:  
3: /cgn2\_6/ptodata/2/pubpna/US06\_NEW\_PUB.seq:  
4: /cgn2\_6/ptodata/2/pubpna/US06\_PUBCOMB.seq:  
5: /cgn2\_6/ptodata/2/pubpna/US07\_NEW\_PUB.seq:  
6: /cgn2\_6/ptodata/2/pubpna/PCTUS\_PUBCOMB.seq:  
7: /cgn2\_6/ptodata/2/pubpna/US08\_NEW\_PUB.seq:  
8: /cgn2\_6/ptodata/2/pubpna/US08\_PUBCOMB.seq:  
9: /cgn2\_6/ptodata/2/pubpna/US09A\_PUBCOMB.seq:  
10: /cgn2\_6/ptodata/2/pubpna/US09B\_PUBCOMB.seq:  
11: /cgn2\_6/ptodata/2/pubpna/US09C\_PUBCOMB.seq:  
12: /cgn2\_6/ptodata/2/pubpna/US09\_NEW\_PUB.seq:  
13: /cgn2\_6/ptodata/2/pubpna/US09\_NEW\_PUB.seq2:  
14: /cgn2\_6/ptodata/2/pubpna/US10A\_PUBCOMB.seq:  
15: /cgn2\_6/ptodata/2/pubpna/US10B\_PUBCOMB.seq:  
16: /cgn2\_6/ptodata/2/pubpna/US10C\_PUBCOMB.seq:  
17: /cgn2\_6/ptodata/2/pubpna/US10\_NEW\_PUB.seq:  
18: /cgn2\_6/ptodata/2/pubpna/US60\_NEW\_PUB.seq:  
19: /cgn2\_6/ptodata/2/pubpna/US60\_PUBCOMB.seq:

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
------------	-------	-------------	--------	----	-------------

ALIGNMENTS

RESULT 1

US-10-045-815-3  
; Sequence 3, Application US/10045815  
; Publication No. US20020160498A1  
; GENERAL INFORMATION:  
; APPLICANT: Madhwa, Renu  
; APPLICANT: Sugihara, Takashi  
; APPLICANT: Ohide, Akiko  
; TITLE OF INVENTION: TUMOR SUPPRESSOR GENE  
; FILE REFERENCE: 06501-091001  
; CURRENT APPLICATION NUMBER: US/10/045,815  
; CURRENT FILING DATE: 2001-10-26  
; PRIOR APPLICATION NUMBER: PCT/JP00/02731  
; PRIOR FILING DATE: 2000-04-26  
; PRIOR APPLICATION NUMBER: JP 11/118806  
; PRIOR FILING DATE: 1999-04-26  
; NUMBER OF SEQ IDS: 15  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 3  
; LENGTH: 2600  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
; FEATURE:

Sequence 3, Appli  
Sequence 228, App  
Sequence 2, Appli  
Sequence 69, Appl  
Sequence 28, Appl  
Sequence 51, Appl  
Sequence 1, Appli  
Sequence 1, Appli  
Sequence 227, App  
Sequence 226, App  
Sequence 6419, Ap  
Sequence 6419, Ap  
Sequence 7, Appli  
Sequence 5, Appli  
Sequence 2327, A  
Sequence 574, App  
Sequence 708, App  
Sequence 708, App  
Sequence 8286, Ap  
Sequence 15266, A  
Sequence 2337, Ap  
Sequence 2337, Ap  
Sequence 3847, Ap  
Sequence 11074, A  
Sequence 13025, A  
Sequence 1401, Ap  
Sequence 123, App  
Sequence 508, App  
Sequence 1793, Ap  
Sequence 217, App  
Sequence 86, Appl  
Sequence 921, App  
Sequence 4475, Ap  
Sequence 618, App  
Sequence 279034,  
Sequence 279035,  
Sequence 279036,  
Sequence 279037,  
Sequence 279034,  
Sequence 279033,  
Sequence 279036,  
Sequence 279037,  
Sequence 8942, Ap  
Sequence 158857,  
Sequence 158859,



NAME/KEY: CDS  
LOCATION: (52)... (2259)  
US-10-045-815-3

## Alignment Scores:

Pred. No.: 0 Length: 2600  
Score: 736.00 Matches: 736  
Percent Similarity: 100.00% Conservativeness: 0  
Best Local Similarity: 100.00% Mismatches: 0  
Query Match: 100.00% Indels: 0  
DB: 14 Gaps: 0

US-10-045-815-4 (1-736) x US-10-045-815-3 (1-2600)

QY 1 MetAlaValArgAlaLeuLysLeuLeuThrThrLeuLeuAlaValAlaAlaSer 20  
DB 52 ATGGCGGTACGCGGCTTGAAGCTGCTGACCACTGCTGGCTGCTGGCGGCTGCCCTCC 111  
QY 21 GlnAlaGluValGluSerGluAlaGlyTyrGlyMetValThrProAspLeuLeuPheAla 40  
DB 112 CAAGCCGAGGTCGAGTCCGAGCGAGGATGGGCGATGGTGCAGCGCTGATCTGCTTCGCC 171  
QY 41 GlnGlyThrAlaAlaTyrAlaArgGlyAspTyrProGlyValValLeuSerMetGluArg 60  
DB 172 GAGGGGACCGGAGCTTACGCGCGGGGACTGGCGCGGGGTGGTCTTGAGCATGGAGACGG 231  
QY 61 AlaLeuArgSerArgAlaAlaLeuArgAlaLeuArgLeuArgCysArgThrGlnCysAla 80  
DB 232 CGCTGCGCTCCGGGCGAGCCCTCCGGGCCCTTCGCTGCTGCCGCCAGTGTGCC 291  
QY 81 AlaAspPheProTyrGluLeuAspProAspTyrSerProSerProAlaGlnAlaSerGly 100  
DB 292 CGCGACTTCCGCTGGGAGCTGGACCCCGACTGCTGCCCGCCAGCGCGGCCAGCGGCTCGGC 351  
QY 101 AlaGlyAlaLeuArgAspLeuSerPheGlyClyLeuLeuArgAlaAlaCysLeu 120  
DB 352 CGCGCGCCCTCGCGGACCTGAGCTTCTTCGGGGGCGCTTCGCGTCCGCTGCGCTGCGCTG 411  
QY 121 ArgArgCysLeuGlyProProAlaAlaHisSerLeuSerGluLeuMetGluLeuGluPhe 140  
DB 412 CGCGCTGCTCGCGCGCGCGCGCGCCACTCGCTCGCTCAGCGAGAGATGGAGTGGAGTTC 471  
QY 141 ArgLysArgSerProTyrAsnTyrLeuGlnValAlaTyrPheLysIleAsnLysLeuGlu 160  
DB 472 CGCAAGCGGAGCCCTACAACTACTCGAGGTCCGCTACTTCAAGATCAACAAGTTGGAG 531  
QY 161 LysAlaValAlaAlaAlaHisThrPhePheValGlyAsnProGluHisMetGluMetGln 180  
DB 532 AAAGCTGTGTGTCAGCACACACCTTCTTCGTGGGCAATCCTGAGCACATGGAAATGCAG 591  
QY 181 GlnAsnLeuAspTyrTyrGlnThrMetSerGlyValLysGluAlaAspPheLysAspLeu 200  
DB 592 CAGAACCTTAGACTATTACCAACCATGCTCGAGGTGAAGGAGCGCGACTTCAAGGATCTT 651  
QY 201 GluThrGlnProHisMetGlnGluPheArgLeuGlyValArgLeuTyrSerGluGluGln 220  
DB 652 GAGACTCAACCCCATATGCAAGANTTTCGACTGGAGTGGGACTCTACTCTCAGAGGACAG 711  
QY 221 ProGlnGluAlaValProHisLeuGluAlaAlaLeuGlnGluTyrPheValAlaTyrGlu 240  
DB 712 CCACAGGAAGTGTGGCCCACTAGAGCGCGCGCTGCAAGATATCTTGTGGGCTATGAG 771  
QY 241 GluCysArgAlaLeuCysGluGlyProTyrAspTyrAspGlyTyrAsnTyrLeuGluTyr 260  
DB 772 GAGTGGCGTCCCTCTCGAAGGCCCTATGACTACCATGGCTACACTACCTTGAGTAC 831  
QY 261 AsnAlaAspLeuPheGlnAlaIleThrAspHisTyrIleGlnValLeuAsnCysLysGln 280  
DB 832 AACGCTGACCTCTTCCAGGCCATCACAGATCATATACATCCAGGTCTCAACTGTAAGCAG 891  
QY 281 AsnCysValThrGluLeuAlaSerHisProSerArgLysProPheGluAspPheLeu 300  
DB 892 AACTGTGTACGGAGCTTGTCTTCCACCCCAAGTCGAGAGAGCCCTTTGAAGACTTCCCTC 951

QY 301 ProSerHisTyrAsnTyrLeuGlnPheAlaTyrTyrAsnIleGlyAsnTyrThrGlnAla 320  
DB 952 CCATCGCATTAATAATATCTGCAGTTTGCTACTATAACATTTGGGAATATATACCAAGCT 1011  
QY 321 GlyGluCysAlaLysThrTyrLeuLeuPhePheProAsnAspGluValMetAsnGlnAsn 340  
DB 1012 GGTGAATGTGCCAGACCTTCTTCTTCCCAATGACGAGGTGATGAACCAAT 1071  
QY 341 LeuAlaTyrTyrAlaAlaMetLeuGlyGluGluHisThrArgSerIleGlyProArgGlu 360  
DB 1072 TTGGGCTATTATGCAGTATGCTTGGAGAAGAACACACAGATCCATCGGCCCGCTGAG 1131  
QY 361 SerAlaLysGluTyrArgGluArgSerLeuLeuLysGluLeuLeuPhePheAlaTyr 380  
DB 1132 AGTGCCAGGAGTACCGACAGCAAGCCTCTGGAAGAAGAACTGCTTTTCTTGGCTTAT 1191  
QY 381 AspValPheGlyIleProPheValAspProAspSerTyrProGluGluValIlePro 400  
DB 1192 GATGTTTTTGGAAATCCCTTTTGGATCCGGATTCATGGACTCCAGAAAGAAAGTATTCCC 1251  
QY 401 LysArgLeuGlnGluLysGlnLysSerGluArgGluThrAlaValArgIleSerGlnGlu 420  
DB 1252 AAGAGATTGCAAGAGAAACAGAAAGTCAGAACGGGAAACAGCGGTACGCATCTCCAGGAG 1311  
QY 421 IleGlyAsnLeuMetLysGluIleGluThrLeuValGluGluLysThrLysGluSerLeu 440  
DB 1312 ATTGGGAACCTTATGAAGAAATCGAGACCTTGTGGAAGAGAGACCAAGAGTCACTG 1371  
QY 441 AspValSerArgLeuThrArgGluGlyGlyProLeuLeuTyrGluGlyIleSerLeuThr 460  
DB 1372 GATGTGAGCAGACTACCCGGGAAGGTGGCCCCCTGCTGTATGAAGGCATCAGTCTCACC 1431  
QY 461 MetAsnSerLysLeuLeuAsnGlyTyrGlnArgValValMetAspGlyValIleSerAsp 480  
DB 1432 ATGAACTCCAACTCTGAAATGGTTACCGAGGGGTGGTGTGATGGAGCGGCTAATCTCTGAC 1491  
QY 481 HisGluCysGlnGluLeuGlnArgLeuThrAsnValAlaAlaThrSerGlyAspGlyTyr 500  
DB 1492 CACGAGTGTGAGGAGCTCGAGAGCTGACCAATGTGGCAGCAACCTCAGGAGATGGGTAC 1551  
QY 501 ArgGlyGlnThrSerProHisThrProAsnGluLysPheTyrGlyValThrValPheLys 520  
DB 1552 CGGGTTCAGACTCCCCACATCTCCCAATGAAGAATCTATGTGTGTCACGTCTTCAA 1611  
QY 521 AlaLeuLysLeuGlyGlnGluLysValProLeuGlnSerAlaHisLeuTyrTyrAsn 540  
DB 1612 GCCCTCAAGCTGGGGCAAGAGGCAAAAGTTCTCTGCAGAGTGGCCACCTGTACTACAC 1671  
QY 541 ValThrGluLysValArgAlaIleMetGluSerTyrPheArgLeuAspThrProLeuTyr 560  
DB 1672 GTGCGGAGAAGTGTGGCGGATCATGAGTCTTCTTCCGCTGGATACGCCCTCTAC 1731  
QY 561 PheSerTyrSerHisLeuValCysArgThrAlaIleGluGluValGlnAlaGluArgLys 580  
DB 1732 TTTTCTTCTCTCTCATCTGTGTGCGGCACTGCCATCGAAGAGGTCCAGGAGAGAGAAAG 1791  
QY 581 AspAspSerHisProValHisValAspAsnCysIleLeuAsnAlaGluThrLeuValCys 600  
DB 1792 GATGATAGTCTATCCAGTCCAGTCCAGCAACTGCAATCTGGAATGCCGAGACCTCGTGTGT 1851  
QY 601 ValLysGluProProAlaTyrThrPheArgAspTyrSerAlaIleLeuTyrLeuAsnGly 620  
DB 1852 GTCAAGAGAGCCCCAGCCCTACACCTTCGCGACTACAGCGCATCTTACTTAATGGG 1911  
QY 621 AspPheAspGlyGlyAsnPheTyrPheThrGluLeuAspAlaLysThrValThrAlaGlu 640  
DB 1912 GACTTCGATGGCGGAAACTTTTATTTCATCTGAATGTCGAAGACCGTGCAGCGAGAG 1971  
QY 641 ValGlnProGlnCysGlyArgAlaValGlyPheSerSerGlyThrGluAsnProHisGly 660  
DB 1972 GTGCAAGCTCTAGTGTGGAGAGCGGTGGGATTTCTTTCAGGCACTGAAACCCACATGGA 2031

QY 661 ValLysAlaValThrArgGlyGlnArgCysAlaIleAlaLeuTrpPheThrLeuAspPro 680  
Db 2032 GTAAAGCGTGTACCAAGGGGAGCGCTGTGCCATCGCCTGTGGTTACCTTGACCCCT 2091  
QY 681 ArgHisSerGluArgAspArgValGlnAlaAspAspLeuValLysMetLeuPheSerPro 700  
Db 2092 CGACACAGCGAGCGGACAGCGGTGTCAGGCGAGATGACCTGGTGAAGATGCTCTTCAGCCCA 2151  
QY 701 GluGluMetAspLeuSerGlnGlnGlnProLeuAspAlaGlnGlnClyProProGluPro 720  
Db 2152 GAAGAGATGGACCTCTCCAGGAGCAGCCCTGGATGCCAGCGGGCCCCCGGAACCT 2211  
QY 721 AlaGlnGluSerLeuSerGlySerGluSerLysProLysAspGluLeu 736  
Db 2212 GCACAGAGTCTCTCTCAGGCACTGATCGAAGCCCCAAGGATGAGCTA 2259

## RESULT 2

US-10-302-172-228  
; Sequence 228, Application US/10302172  
; Publication No. US20040053250A1  
; GENERAL INFORMATION:  
; APPLICANT: Tang, Y. Tom  
; APPLICANT: Xue, Aidong J.  
; APPLICANT: Drmanac, Radoje T.  
; TITLE OF INVENTION: No. US20040053250A1 Arginine-rich Protein-like Nucleic Acids and  
; TITLE OF INVENTION: Polypeptides  
; FILE REFERENCE: 803.1CNCIP  
; CURRENT APPLICATION NUMBER: US/10/302,172  
; CURRENT FILING DATE: 2002-11-21  
; PRIOR APPLICATION NUMBER: US 10/235,251  
; PRIOR FILING DATE: 2002-08-20  
; PRIOR APPLICATION NUMBER: PCT US02/05095  
; PRIOR FILING DATE: 2002-03-05  
; PRIOR APPLICATION NUMBER: US 09/799,451  
; PRIOR FILING DATE: 2001-03-05  
; NUMBER OF SEQ ID NOS: 950  
; SOFTWARE: pt\_FL\_genes Version 2.0  
; SEQ ID NO 228  
; LENGTH: 2242  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
; FEATURE:  
; NAME/KEY: CDS  
; LOCATION: (32)..(2239)  
US-10-302-172-228

Alignment Scores:  
Pred. No.: 0 Length: 2242  
Score: 634.00 Matches: 734  
Percent Similarity: 99.46% Conservative: 0  
Best Local Similarity: 99.46% Mismatches: 2  
Query Match: 86.14% Indels: 4  
DB: 13 Gaps: 0

US-10-045-815-4 (1-736) x US-10-302-172-228 (1-2242)

QY 1 MetAlaValArgAlaLeuLysLeuLeuThrThrLeuLeuAlaValAlaAlaLaser 20  
Db 32 ATGGCGGTACCGCGTTCAAGCTGCTGACCACACTGTGCTGGCTGTGCGGCGCTGCTCC 91  
QY 21 GlnAlaGluValGluSerGluAlaGlyTrpGlyMetValThrProAspLeuPheAla 40  
Db 92 CAAGCCGAGGTGAGTTCGAGGAGAGATGGGGCATGGTGACGCTGATCTGCTCTTCGCC 151  
QY 41 GluGlyThrAlaAlaTyAla-ArgGlyAspTrpProGlyValValLysSerMetGluAr 60  
Db 152 GAGCGGACCGCAGCGCTACTC-GCGCGGGAGCTGGCCCGGGGTGGTCTCTGAGCATGGAACG 210  
QY 60 gAlaLeuArgSerArgAlaAlaLeuArgAlaLeuArgLeuArgCysArgThrGlnCysAl 80  
Db 211 GGCGCTGCGCTCCCGGGGAGCCCTCCGCGCCCTTCGCTGGCGTGGCCGCCACCGAGTGGC 270  
QY 80 aAlaAspPheProTrpTrpGluLeuAspProAspTrpSerProAlaGlnAlaSerG1 100

Db 271 CGCGGACTTCCGTGGGAGCTGGACCCGACCTGGTCCCGCCAGCGCGCCAGGCGCTCGG 330  
QY 100 yAlaGly-AlaLeuArgAspLeuSerPhePheGlyGlyLeuLeuArgArgAlaAlaCysL 120  
Db 331 CGCGCGC-CGCGCTCGGAGACCTGAGCTTCTTGGGGGGCCCTTCTGCTGCGCGCTCGCTGCC 389  
QY 120 euArgArgCysLeuGlyProProAlaAlaHisSerLeuSerGluGluMetGluLeuGluP 140  
Db 390 TGGCGCGCTGGCTCGGCGCGCGCGCGCCCACTCGCTCAGCGAAGAGATGGAGCTGGAGT 449  
QY 140 heArgLysArgSerProTrpTrpAsnTrpLeuGlnValAlaTyPheLysLysLeuLysLeuG 160  
Db 450 TCCGCAAGCGGAGCGCTTACAACTACTCTGAGTGGCGCTACTTCAAGATCAACAAGTTGG 509  
QY 160 luLysAlaValAlaAlaAlaHisThrPhePheValGlyVAsnProGluHisMetGluMetG 180  
Db 510 AGAAAGCTGTGTGTCAGCACACACCTTCTTCTGGGGGCAATCCTGAGCACATGGNAATGC 569  
QY 180 lnGlnAsnLeuAspTrpTrpGlnThrMetSerGlyValLysGluAlaAspPheLysAspL 200  
Db 570 AGCAGAACCTAGACTATTACCAACCATGTCTGGAGTGAAGGAGCGCCACTTCAAGGATC 629  
QY 200 euGluThrGlnProHisMetGlnGluPheArgLeuGlyValArgLeuTrpSerGluGluG 220  
Db 630 TTGAGACTCAACCCCATATGCAAGAAATTCGACTGGGAGTGGCACTTACTCAGAGGAAC 689  
QY 220 lnProGlnGluAlaValProHisLeuGluAlaAlaLeuGlnGluTrpPheValAlaTyR 240  
Db 690 AGCCACAGAGAGCTGTGCCCCACCTAGAGGCGCGCTGCAAGAATACITTTGTGGCTATG 749  
QY 240 luGlyCysArgAlaLeuCysGluGlyProTrpAspTrpAspGlyTyRAsnTrpLeuGluT 260  
Db 750 AGGAGTGGCGTGGCTCTGCGAAGGGCCCTATGACTACGATGGCTACCACTACCTTGAGT 809  
QY 260 yRAsnAlaAspLeuPheGlnAlaIleThrAspHisTyRileGlnValLeuAsnCysLysG 280  
Db 810 ACAACGCTGACCTTCCAGGCCATCACAGATCATATCATCCAGTCTCTCACTGTAAAGC 869  
QY 280 lnAsnCysValThrGluLeuAlaSerHisProSerArgGluLysProPheGluAspPheL 300  
Db 870 AGAAGCTGTGTACGGAGCTTGTCTCCACCCCAAGTCGAGAGAAAGCCCTTGAAGACTTCC 929  
QY 300 euProSerHisTyRAsnTrpLeuGlnPheAlaTyRAsnIleGlyVAsnTrpThrGlnA 320  
Db 930 TCCCATCGCATTAATAATTATCTGCAGTTGCTTACTATAACATTGGGAATTATACACAG 989  
QY 320 laGlyGluCysAlaLysThrTyRLeuLeuPhePheProAsnAspGluValMetAsnGlnA 340  
Db 990 CTGGTGAAATGTCGCAAGACCTATCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTT 1049  
QY 340 snLeuAlaTyRTrpAlaAlaMetLeuGlyGluGluHisThrArgSerIleGlyProArgG 360  
Db 1050 ATTTGGCTTATTATGTCAGCTATGCTTGGAGAAGAAACACACAGATCCATCGGCCCTCGTG 1109  
QY 360 luSerAlaLysGluTrpArgGlnArgSerLeuLeuGluLysGluLeuLeuPhePheAlaT 380  
Db 1110 AGAGTCCCAAGAGTACCGACAGCGAGCCTACTCTGAAAAAAGAACTGCTTTTCTTCTGCTT 1169  
QY 380 yRAspValPheGlyIlePhePheValAspProAspSerTrpThrProGluGluValIleP 400  
Db 1170 ATGATGTTTTTGGAAATTCCTTTTGGATCCGGATTCATGGACTCCAGAAAGATGATTC 1229  
QY 400 rplysArgLeuGlnGluLysGlnLysSerGluArgGluThrAlaValArgIleSerGlnG 420  
Db 1230 CCAAGAGATTGCAAGAGAAACAGAAAGTCAAGACGGGAAACACCGCTTACCATCTCCAGG 1289  
QY 420 luIleGlyAsnLeuMetLysGluIleGluThrLeuValGluGluLysThrLysGluSerL 440  
Db 1290 AGATTGGGAACCTTATGAAGGAAATCGAGACCCCTTGTGGAGAGAAGAACCAAGGAGTCC 1349  
QY 440 euAspValSerArgLeuThrArgGluGlyGlyProLeuLeuTrpGluGlyIleSerLeuT 460

```
Db 1350 TCGATGTGAGCAGACTGACCGGAGGTGGCCCTGCTGTATGAGGCACTAGTCTCA 1409
Qy hrMetAsnSerLysLeuLeuAsnGlyTyrGlnArgValValMetAspGlyValIleSerA 480
Db 1410 CCATGAACCTCCAACTCTGAATGTTACCAAGCGGTGGTGATGACGGGTAACTCTCTG 1469
Qy spHisGluCysGlnGluLeuGlnArgLeuThrAsnValAlaAlaThrSerLysAspGlyT 500
Db 1470 ACCAGAGTGTGAGAGCTGCAGAGACTGACCAATGTGGCAGCAACTCTCAGGAGTGGCT 1529
Qy yrArgGlyGlnThrSerProHisThrProAsnGluLysPheTyrGlyValThrValPheL 520
Db 1530 ACCGGGTGCAGACCTCCCACTACTCCCAATGAAGATTCTATGGTGTCACTGTCTTCA 1589
Qy ysAlaLeuLysLeuGlyGlnGluLysValProLeuGlnSerAlaHisLeuTyrTyrA 540
Db 1590 AAGCCCTCAAGCTGGGCAAGAGGCAAAAGTTCCTCTGCAGAGTGGCCACCTGTACTACA 1649
Qy snValThrGluLysValArgArgIleMetGluSerTyrPheArgLeuAspThrProLeuT 560
Db 1650 ACGTACGGAGAAAGTGGCGGCATCATGGAGTCTACTCCGCTCGATACGCCCTCT 1709
Qy yrPheSerTyrSerHisLeuValCysArgThrAlaIleGluGluValGluAlaGluArgL 580
Db 1710 ACTTTTCTACTCTCATCTGTGTGGCGCACTGCATCGAAGAGGTCCAGGCAAGAGGA 1769
Qy ysAspAspSerHisProValHisValAspAsnCysIleLeuAsnAlaGluThrLeuValC 600
Db 1770 AGGATGATAGTATCCAGTCCAGTGCAGCAACTGCATCTTGAATGCCGAGACCTGTGT 1829
Qy ysValLysGluProProAlaTyrThrPheArgAspTyrSerAlaIleLeuTyrLeuAsnG 620
Db 1830 GTGTCAAGAGAGCCCAAGCTTACCTTCGCGACTACAGCGCCATCTTTACCTAAATG 1889
Qy lyAspPheAspGlyCysAsnPheTyrPheThrGluLeuAspAlaLysThrValThAlaG 640
Db 1890 GGGACTTCGATGGCGGAACTTTATTTCATCTGAATGGATGCCAGACCGTACGCGCAG 1949
Qy luValGlnProGlnCysGlyArgAlaValGlyPheSerSerGlyThrGluAsnProHisG 660
Db 1950 AGGTGACAGCTCAGTGTGGAAGACCGTGGGATTCCTTCAGGCACTGAAACCCACATG 2009
Qy lyValLysAlaValThrArgGlyGlnArgCysAlaIleAlaLeuThrPheThrLeuAspP 680
Db 2010 GAGTGAAGGTGTACCAAGGGGAGCGCTGTGCGCATCGCCCTGTGCTTACCCCTGGACC 2069
Qy roAtsHisSerGluArgAspArgValGlnAlaAspAspLeuValLysMetLeuPheSerP 700
Db 2070 CTCGACACAGCGAGCGGACAGGAGTGCAGGAGATGACCTGGTGAAGATGCTCTTCAGCC 2129
Qy roGluGluMetAspLeuSerGlnGluGlnProLeuAspAlaGlnGlnGlyProProGluP 720
Db 2130 CAGAAGAGATGGACCTCTCCAGAGAGAGCGCCCTGGATGCCAGAGGGCCCCCGGAAAC 2189
Qy roAlaGlnGluSerLeuSerGlySerGluSerLysProLysAspGluLeu 736
Db 2190 CTGCACAGAGTCTCTCTCAGGAGTGAATCGAAGCCCAAGGATGAGCTA 2239
```

## RESULT 3

```
US-10-257-174-2
; Sequence 2, Application US/10257174
; Publication No. US20040034194A1
; GENERAL INFORMATION:
; APPLICANT: Agarwal, Pankaj
; APPLICANT: Murdock, Paul R.
; APPLICANT: Rizvi, Safia K.
; APPLICANT: Smith, Randall F.
; APPLICANT: Xiang, Zhaoying
; TITLE OF INVENTION: NOVEL COMPOUNDS
; FILE REFERENCE: GP50022
; CURRENT APPLICATION NUMBER: US/10/257,174
; CURRENT FILING DATE: 2002-10-10
; PRIOR APPLICATION NUMBER: PCT/US01/11797
```

```
; PRIOR FILING DATE: 2001-04-11
; PRIOR APPLICATION NUMBER: 60/196,603
; PRIOR FILING DATE: 2000-04-13
; PRIOR APPLICATION NUMBER: 60/199,417
; PRIOR FILING DATE: 2000-04-24
; NUMBER OF SEQ ID NOS: 48
; SOFTWARE: FastSeq for Windows version 3.0
; SEQ ID NO 2
; LENGTH: 2211
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-257-174-2
```

## Alignment Scores:

```
Pred. No.: 0 Length: 2211
Score: 433.00 Matches: 733
Percent Similarity: 99.19% Conservative: 0
Best Local Similarity: 99.19% Mismatches: 3
Query Match: 58.83% Indels: 6
DB: 13 Gaps: 0
```

US-10-045-815-4 (1-736) x US-10-257-174-2 (1-2211)

```
Qy 1 MetAlaValArgAlaLeuLysLeuLeuThrThrLeuLeuAlaValAlaAlaAlaSer 20
Db 1 ATGGCGGTACGCGCTTGAAGCTGCTGACACACACTGCTGGCTGTCTGGCGCTGCTCC 60
Qy 21 GlnAlaGluValGluSerGluAlaGlyTyrGlyMetValThrProAspLeuLeuPheAla 40
Db 61 CAAGCCGAGGTGAGTCCGAGCAGGATGGGCGCATGGTACGCTGATCTCTCTCGCC 120
Qy 41 GluGlyThrAlaAlaTyrAlaArgGlyAspTrpProGlyValValLeuSerMetGluArg 60
Db 121 GAGGGGACCGCAGCTACGCGCGGGGACTGGCCCGGGTGGTCTTGAGCATGGAAAG 180
Qy 61 AlaLeuArgSerArgAlaAlaLeuArgAlaLeuArgLeuArgCysArgThrGlnCysAla 80
Db 181 GCGCTGCGCTCCCGGCGACGCTCCGCGCCCTTGGCTGCGCTGCGCGCACCGAGTGGCC 240
Qy 81 AlaAspPheProTyrGluLeuAspProAspTrpSerProSerProAlaGlnAlaSerGly 100
Db 241 GCGGACTTCCCGTGGAGCTGGAACCCGACTGTGTCCCGCCGCGCCGCGCTCGGGC 300
Qy 101 AlaGly-AlaLeuArgAspLeuSerPhePheGlyClyLeuLeuArgArgAlaAlaCysLe 120
Db 301 GCGGC- CGCCCTCGCGCACCTGAGCTTCTTCGGGGGCTTCTGCGTGGCTGCTGCT 359
Qy 120 uArgArgCysLeuGlyProProAlaAlaHisSerLeuSerGluMetGluLeuGluPh 140
Db 360 GCGCGCTGCTTCGGGCGCGCGCCCTGCTGCTGAGGAGATGGAGCTGGAGTT 419
Qy 140 eArgLysArgSerProTyrAsnTyrLeuGlnValAlaTyrPheLysIleAsnLysLeuGl 160
Db 420 CGCAAGCGGAGCGCTTACCACTACCTGCTGCTGCTGCTTCAAGATCAACAGTTGGA 479
Qy 160 uLysAlaValAlaAlaAlaHisThrPhePheValGlyAsnProGluHisMetGluMetGl 180
Db 480 GAAAGCTTGTTCGAGCACACACTTCTTCGTGGGCAATCTCTGAGCACATGGAAATGCA 539
Qy 180 nGlnAsnLeuAspTyrTyrGlnThrMetSerGlyValLysGluAlaAspPheLysAspLe 200
Db 540 GCAGAACCTAGACTATTACCAACCATGTCTGGAGTGAAGAGGCGCCACTTCAAGGATCT 599
Qy 200 uGluThrGlnProHisMetGlnGluPheArgLeuGlyValArgLeuTyrSerGluGluGl 220
Db 600 TGAGACTCACCCCATATGCAAGAAATTTCCGACTGGGAGTGGCACTCTACTCAGAGGAACA 659
Qy 220 nProGlnGluAlaValProHisLeuGluAlaAlaLeuGlnGluTyrPheValAlaTyrGl 240
Db 660 GCCACAGGAAGCTGTGCCCCACCTAGAGGGCGCTGCAAGATATCTTTGGGCTATGA 719
Qy 240 uGluCysArgAlaLeuCysGluGlyProTyrAspTyrAspGlyTyrAsnTyrLeuGluTyr 260
```

720 GGAGTGGCGTCCCTCTGCGAAGGGCCCTATGACTACGATGGCTACAACACTACCTTGAGTA 779  
Qy rAsnAlaAspLeuPheGlnAlaIleThrAspHisTyrIleGlnValLeuAsnCysLysG1 280  
Db CAACGCTGACCTCTCCAGGCCATCACAGATCATTTACATCCAGTCTCTCAACTGTAAGCA 839  
Qy nAsnCysValThrGluLeuAlaSerHisProSerArgGluLeuPheProPheGluAspPheLe 300  
Db GAACTGTGTCCAGGAGCTTCTTCCACCCCAAGTCGAGAGAGCCCTTTGAAGACTTCT 899  
Qy uProSerHisTyrAsnTyrLeuGlnPheAlaTyrTyrAsnIleGlyAsnTyrThrGlnAl 320  
Db CCCATGCCATTATATCTGAGTTTGCCTACTATATACATTTGGGAATATACACAGGC 959  
Qy aGly-GluCysAlaLysThrTyrLeuLeuPhePheProAsnAspGluValMetAsnGlnA 340  
Db TGT-TGAATGTGCAAGACCTATCTCTCTTCCCAATGACGAGGTGATGAACCAA 1018  
Qy snLeuAlaTyrTyrAlaAlaMetLeuGlyGluHisThrArgSerIleGlyProArgG 360  
Db ATTGGCCTATTATGCACTATCTTGGAGAGAACACACAGATCCATCGGCCCCCGTG 1078  
Qy lueSerAlaLysGluTyrArgGlnArgSerLeuLeuGluLysGluLeuLeuPhePheAla 380  
Db AGATGCCAAGGAGTACCGACAGCGAGCCCTACTGGAAAAGAACTGCTTTTCTCGCTT 1138  
Qy YrAspValPheGlyIleProPheValAspProAspSerThrThrProGluGluValIleP 400  
Db ATGATGTTTTTGGAAATTCCTTTGTGGATCCGATTCATGGACTCCAGAGAGTGAATTC 1198  
Qy rOlyArgLeuGlnGluLysGlnLysSerGluArgGluThrAlaValArgLysSerGlnG 420  
Db CCAAGAGATTGCAAGAGAACAGAGTCAAGACGGGAAACAGCCGTACCGATCTCCAGG 1258  
Qy lueGlyAsnLeuMetLysGluLeuThrLeuValGluGluLysThrLysGluSerL 440  
Db AGATTGGAACTTATGAAGAAATCGAGACCTTGTGGAGAGAGAACCAAGAGTCA 1318  
Qy euAspValSerArgLeuThrArgGluGlyProLeuLeuTyrGluGlyLysSerLeuT 460  
Db TGGATGTGACGAGACTGACCCGGAGGTGGCCCTCTGTGTATGAGGATCATGCTCA 1378  
Qy hrMetAsnSerLysLeuLeuAsnGlyTyr-GlnArgValValMetAspGlyValIleSer 479  
Db CCATGAACTCCAACTCTCTGAATGGTTC-CCAGCGGTGGTGTATGAGCGCGTAATCTCT 1437  
Qy AspHisGluCysGlnGluLeuGlnArgLeuThrAsnValAlaThrSerGlyAspGly 499  
Db GACCACAGTGTCCAGGAGCTGCAGAGACTGACCAATGTGGCAACCACTCAGAGATGGC 1497  
Qy TyrArgGlyGlnThrSerProHisThrProAsnGluLysPheTyrGlyValThrValPhe 519  
Db TACCGGGGTGAGACTCCCCACATCTCCCAATGAAAGTTCTATGGTGTCACTGTCTTC 1557  
Qy LysAlaLeuLysLeuGlnGluGlyLysValProLeuGlnSerAlaHisLeuTyrTyr 539  
Db AAAGCCCTCAAGCTGGGGCAAGGCAAGGAAAGTTCTCTGAGAGTGCACCTGTACTAC 1617  
Qy AsnValThrGluLysValArgAlaGlyIleMetGluSerTyrPheArgLeuAspThrProLeu 559  
Db AACGTGACGAGAGGGTGGCGGATCATGAGTGTCTTCCGCTGGATAGCCCTC 1677  
Qy TyrPheSerTyrSerHisLeuValCysArgThrAlaIleGluGluValGlnAlaGluArg 579  
Db TACTTTTCTCTCTCTCTGTTGTCGGCACTGCCATCGAAGAGTCCAGGCAGAGAGG 1737  
Qy LysAspAspSerHisProValHisValAspAsnCysIleLeuAsnAlaGluThrLeuVal 599  
Db AAGGATGATGATCATCAGTCCAGTCCAGTCACTGCACTCTGTAATGCCGAGCCCTCGT 1797  
Qy CysValLysGluProProAlaTyrThrPheArgAspTyrSerAlaIleLeuTyrLeuAsn 619  
Db TGTGTCAAGAGCCCCCAGCTACACCTTCGGGACTACAGCGCATCTTACCTAAAT 1857

Qy GlyAspPheAspGlyGlyAsnPheTyrPheThrGluLeuAspAlaLysThrValThrAla 639  
Db GGGACTTCGATGGCGGAAACTTTTATTTCACTGAATGGATGCCAAGACCGTACGGCA 1917  
Qy GluValGlnProGlnCysGlyArgAlaValGlyPheSerSerGlyThrGluAsnProHis 659  
Db GAGTGTGACCTCAGTGTGGAGAGCCGTGGATTTCTTTCAGGCATGAAAACCCACAT 1977  
Qy GlyValLysAlaValThrArgGlyGlnArgCysAlaIleAlaLeuTyrPheThrLeuAsp 679  
Db GGAGTGAAGGTGTCTCACCAGGGGCGAGCGTGTGCTGCTGCTGCTGCTGCTGCTGCTG 2037  
Qy ProArgHisSerGluArgAspArgValGlnAlaAspAspLeuValLysMetLeuPheSer 699  
Db CTTGACACAGCGAGCGGAGGAGGTGCGAGCAGATGACCTGGTGAAGATGCTCTTTCAGC 2097  
Qy ProGluGluMetAspLeuSerGlnGluGlnProLeuAspAlaGlnGlnGlyProProGlu 719  
Db CCAGAGAGATGGACTCTCCAGGAGCAGCCCTGGATGCCAGCAGGGTCCCCCGCAA 2157  
Qy ProAlaGlnGluSerLeuSerGlySerGluSerLysProLysAspGluLeu 736  
Db CCTGCACAGAGTCTCTCTCAGCAGTGAATCGAAGCCCAAGGATGAGCTA 2208

## RESULT 4

US-10-312-352-69  
; Sequence 69, Application US/10312352  
; Publication No. US20040053824A1  
; GENERAL INFORMATION:  
; APPLICANT: INCYTE GENOMICS, INC.; TANG, Y. Tom  
; APPLICANT: YUE, Henry; AZIMZAI, Yalda  
; APPLICANT: HE, Ann; BATRA, Sajeev  
; APPLICANT: LO, Terence P.; NGUYEN, Damien B.  
; APPLICANT: BURRILL, John D.; MARCUS, Gregory A.  
; APPLICANT: ZINGLER, Kurt A.; GANDHI, Ameena R.  
; APPLICANT: LAL, Preeti G.; KEANEY, Liam  
; APPLICANT: BURFORD, Neil; YAO, Monique G.  
; APPLICANT: CHAWLA, Narinder K.; ELLIOT, Vicki S.  
; APPLICANT: ARVIZU, Chandra S.; KHAN, Farrah A.  
; APPLICANT: BAUGHN, Mariyah R.; HAPALIA, April, J.A.  
; APPLICANT: POLICKY, Jennifer L.; AU-YOUNG, Janice K.  
; APPLICANT: LU, Yan; BOROWSKY, Mark L.  
; APPLICANT: LU, Dying Aina M.; RAMKUMAR, Jayalaxmi  
; APPLICANT: YANG, Junming; GURURAJAN, Rajagopal  
; APPLICANT: WARREN, Bridget A.; GIETZEN, Kimberly J.  
; APPLICANT: XU, Yuming; KALLICK, Deborah A. Kavitha  
; APPLICANT: LEE, Ernestine A.; THANGAVELU, Kavitha  
; APPLICANT: DELEGEANE, Angelo M.; LEE, Sally  
; TITLE OF INVENTION: EXTRACELLULAR MATRIX AND CELL ADHESION MOLECULES  
; FILE REFERENCE: PF-0794 USN  
; CURRENT APPLICATION NUMBER: US/10/312,352  
; PRIOR FILING DATE: 2002-12-18  
; PRIOR APPLICATION NUMBER: PCT/US01/21067  
; PRIOR FILING DATE: 2001-08-29  
; PRIOR APPLICATION NUMBER: US 60/215,454  
; PRIOR FILING DATE: 2000-06-30  
; PRIOR APPLICATION NUMBER: US 60/219,462  
; PRIOR FILING DATE: 2000-07-18  
; PRIOR APPLICATION NUMBER: US 60/240,111  
; PRIOR FILING DATE: 2000-10-12  
; PRIOR APPLICATION NUMBER: US 60/240,106  
; PRIOR FILING DATE: 2000-10-12  
; PRIOR APPLICATION NUMBER: US 60/244,021  
; PRIOR FILING DATE: 2000-10-27  
; PRIOR APPLICATION NUMBER: US 60/248,887  
; PRIOR FILING DATE: 2000-11-14  
; PRIOR APPLICATION NUMBER: US 60/249,570  
; PRIOR FILING DATE: 2000-11-16  
; NUMBER OF SEQ ID NOS: 72  
; SOFTWARE: PERL Program  
; SEQ ID NO 69  
; LENGTH: 2583

```

; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc feature
; OTHER INFORMATION: Incyte ID No. US20040053824A1 6780147CB1
US-10-312-352-69

Alignment Scores:
Pred. No.: 0 Length: 2583
Score: 433.00 Matches: 733
Percent Similarity: 99.19% Conservative: 0
Best Local Similarity: 99.19% Mismatches: 3
Query Match: 13 Indels: 6
DB: 13 Gaps: 0

US-10-045-815-4 (1-736) x US-10-312-352-69 (1-2583)

QY 1 MetAlaValArgAlaLeuLysLeuLeuThrThrLeuLeuAlaValAlaAlaSer 20
DB 42 ATGGCGGTACCGCGGTGAAGCTGCTGACCACTGCTGGCTGCTGGCGCGCTGCCCTCC 101
QY 21 GlnAlaGluValGluSerGluAlaGlyTyrPheGlyMetValThrProAspLeuLeuPheAla 40
DB 102 CAAGCCGAGGTTCGAGTCCGAGCGAGGATGGGCGATGGTGCCTGATCTCTTCGCC 161
QY 41 GluGlyThrAlaAlaTyrAlaArgGlyAspTyrProGlyValValLeuSerMetGluArg 60
DB 162 GAGGGGACCGGAGCTACGCGCGCGGGGACTGGCGCGGGTGGTCTTCAGCATGGAAACGG 221
QY 61 AlaLeuArgSerArgAlaAlaLeuArgAlaLeuArgLeuArgCysArgThrGlnCysAla 80
DB 222 GCGTGGCTCCCGGAGAGCTCCGCGCCCTTCGCTGCTGCTGCTGCTGCTGCTGCTGCTG 281
QY 81 AlaAspPheProTyrGluLeuAspProAspTyrSerProSerProAlaGlnAlaSerGly 100
DB 282 GCCGACTTCCCGTGGAGCTGGACCCCGACTGGTCCCGCCGCGCGCGCCAGGCGCTCGGC 341
QY 101 AlaGly-AlaLeuArgAspLeuSerPhePheGlyGlyLeuLeuArgArgAlaAlaCysLe 120
DB 342 GCCGC-CGCCCTGCGGAGCTGAGCTTCTTCGGGGGCTTCGCGTCCGCTGCTGCTGCT 400
QY 120 uArgArgCysLeuGlyProProAlaAlaHisSerLeuSerGluGluMetGluLeuGluPh 140
DB 401 CGCGCGCTGCTCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 460
QY 140 eArgLyArgSerProTyrAsnTyrLeuGlnValAlaTyrPheLysIleAsnLysLeuGlu 160
DB 461 CCGCAAGCGGAGCCCTACAACTACCTGCGAGGTGCGCTTCAAGATCAACAAGTTGGA 520
QY 160 uLysAlaValAlaAlaAlaHisThrPhePheValGlyAsnProGluHisMetGluMetGlu 180
DB 521 GAAAGCTGTGCTGCAGCACACACCTTCTCGTGGGCAATCCTGAGCACATGGAATGCA 580
QY 180 nGlnAsnLeuAspTyrTyrGlnThrMetSerGlyValLysGluAlaAspPheLysAspLe 200
DB 581 GCAGAACTAGACTATTACCAACCACTGTCTGGAGTGAAGAGCGCGGACTTCAAGGATCT 640
QY 200 uGluThrGlnProHisMetGlnGluPheArgLeuGlyValArgLeuTyrSerGluGluGlu 220
DB 641 TGAGACTCAACCCCATATGACAGAAATTCGACTGGGAGTGGCGACTCTACTCAGAGGAACA 700
QY 220 nProGlnGluAlaValProHisLeuGluAlaAlaLeuGlnGluTyrPheValAlaTyrGlu 240
DB 701 GCCACAGGAAGCTGTGCGCCCACTTAGAGCGCGGCTGCAAGAAATACTTTGTGGCTATGA 760
QY 240 uGluCysArgAlaLeuCysGluGlyProTyrAspTyrAspGlyTyrAsnTyrLeuGluTyr 260
DB 761 GGAGTGGCGTCCCTCTGCGAAGGCGCTATGACTAGTAGGTGCTACAACTTACCTTGGAT 820
QY 260 rAsnAlaAspLeuPheGlnAlaLeuThrAspHisTyrIleGlnValLeuAsnCysLysGlu 280
DB 821 CAACGCTGACTCTTCCAGGCGCATCACAGATCATTTACATCCAGTCCCTCACTGTAAAGCA 880
QY 280 nAsnCysValThrGluLeuAlaSerHisProSerArgGluLysProPheGluAspPheLe 300
DB 881 GAACGTGTGTACGGAGCTTGTCCACCCCAAGTCGACAGAAGCCCTTTGAAGACTTCT 940
QY 300 uProSerHisTyrAsnTyrLeuGlnPheAlaTyrTyrAsnIleGlyAsnTyrThrGlnAl 320
DB 941 CCCATCGCATTAATAATATCTGCAGTTTGGCTTACTATAACATTTGGAAATATACACAGGC 1000
QY 320 aGly-GluCysAlaLysThrTyrLeuLeuPhePheProAsnAspGluValMetAsnGlnA 340
DB 1001 TGT-TGAATGTCGCAAGACCTATCTTCTTCTTCCCAANTGACGAGGTGATGAACCAA 1059
QY 340 snLeuAlaTyrTyrAlaAlaMetLeuGlyGluGluHisThrArgSerIleGlyProArgG 360
DB 1060 ATTTGGCTTATTATGCAGCTATGCTTGGAGAAGAACACACACAGATCCATCGCGCCCGCTG 1119
QY 360 lUserAlaLysGluTyrArgGlnArgSerLeuLeuGluLysGluLeuLeuPheAlaT 380
DB 1120 AGAGTCCGAGGAGTACCGACAGCAAGCCCTACTGGAAGAAAGAACTGTTTTCTTCGTT 1179
QY 380 YrAspValPheGlyIleProPheValAspProAspSerTyrThrProGluGluValIleP 400
DB 1180 ATGATGTTTTTGGAAATTCCTTTGTGGATCCGGAATCATGGAATCCAGAGAGTGAITC 1239
QY 400 rLysArgLeuGlnLysGlnLysSerGluArgGluThrAlaValArgIleSerGlnG 420
DB 1240 CCAAGAGATTCGACAGAGAAACAGAGTCAGAACCGGAAACAGCCGTACGCCATCTCCCAAG 1299
QY 420 luIleGlyAsnLeuMetLysGluIleGluThrLeuValGluGluLysThrLysGluSerL 440
DB 1300 AGATTGGGAACCTTATGAAGGAATCGAGACCTTGTGGAAGAGAGACCAAGAGGTAC 1359
QY 440 euAspValSerArgLeuThrArgGluGlyGlyProLeuLeuTyrGluGlyIleSerLeuT 460
DB 1360 TGGATGTGAGCAGACTGACCCGGGAAGGTGCCCTGCTGTATGAAGGCATCAGTCTCA 1419
QY 460 hrMetAsnSerLysLeuLeuAsnGlyTyr-GlnArgValValMetAspGlyValIleSer 479
DB 1420 CCATGAACTCCAACTCTGAAATGGTTC-CCAGCGGGTGGTGTGAGGAGCGGTAACTCT 1478
QY 480 AspHisGluCysGlnGluLeuGlnArgLeuThrAsnValAlaAlaThrSerGlyAspGly 499
DB 1479 CACCAACGAGTGTACGAGCTGCAGAGACTGACCAATGTGGCAGCAACCTCAGGAGATGC 1538
QY 500 TyrArgGlyGlnThrSerProHisThrProAsnGluLysPheTyrGlyValThrValPhe 519
DB 1539 TACCGGGGTGAGACCTCCCCACATCTCCCAATGAAAGTTCTATGGTGTCACTGTCTTC 1598
QY 520 LysAlaLeuLysLeuGlyGlnGluLysValProLeuGlnSerAlaHisLeuTyrTyr 539
DB 1599 AAAGCCCTCAAGCTGGGCAAGAGGCAAAAGTTCCTCTGCAGAGTGGCCACCTGTACTAC 1658
QY 540 AsnValThrGluLysValArgArgIleMetGluSerTyrPheArgLeuAspThrProLeu 559
DB 1659 AACGTGACGGAAGGTGCGCGCATCATGAGTCCCTACTTCCCGCTGGATACGCCCTC 1718
QY 560 TyrPheSerTyrSerHisLeuValCysArgThrAlaIleGluGluValGlnAlaGluArg 579
DB 1719 TACTTTCTCTACTCTCATCTGTTGTGCGCACTGCCATCGAAGAGGTCCAGGAGAGAGG 1778
QY 580 LysAspAspSerHisProValHisValAspAsnCysIleLeuAsnAlaGluThrLeuVal 599
DB 1779 AAGGATGATGATCATCCAGTCCAGTCCAGCAACTGCAATCTCTGAATGCGAGACCTCGTG 1838
QY 600 CysValLysGluProProAlaTyrThrPheArgAspTyrSerAlaIleLeuTyrLeuAsn 619
DB 1839 TGTGTCAAGAGAGCCCGAGCCCTACACCTTCGCGACTACAGCGCCATCTTACTTAAT 1898
QY 620 GlyAspPheAspGlyGlyAsnPheTyrPheThrGluLeuAspAlaLysThrValThAla 639
DB 1899 GGGGACTTCGATGCGGAAACTTTTATTTACTACTGAACTGAACTGCAAGACCGTGACGCA 1958
QY 640 GluValGlnProGlnCysGlyArgAlaValGlyPheSerSerGlyThrGluAsnProHis 659
```

Db 1959 GAGGTGACAGCTCAGTGTGGAAGAGCCGTGGATTCTCTTACGCACTGAAACCCACAT 2018  
Qy 660 GlyValLysAlaValThrArgGlyGlnArgCysAlaIleAlaLeuTrpPheThrLeuAsp 679  
Db 2019 GGAGTGAAGGCTGTCAACAGGGGGCAGCGCTGTGCCATCGCCCTGTGGTTACCCCTGGAC 2078  
Qy 680 ProArgHisSerGluArgAspArgValGlnAlaAspLeuValLysValLysPheSer 699  
Db 2079 CCTCGACACAGCAGCGGACAGGTCAGCGACAGATGACCTGTGAAGATGCTCTTCAGC 2138  
Qy 700 ProGluGluMetAspLeuSerGlnGlnProLeuAspAlaGlnGlnGlyProProGlu 719  
Db 2139 CCAGAGAGATGACCTCTCCAGGAGCAGCCCTGGATGCCAGCAGGCGCCCGCGAA 2198  
Qy 720 ProLacGlnGluSerLeuSerGlySerGluSerLysProLysAspGluLeu 736  
Db 2199 CCTGCACAAGAGTCTCTCTCAGGCAGTGAATCGAAGCCCAAGGATGAGCTA 2249

## RESULT 5

US-09-728-952-28  
; Sequence 28 Application US/09728952  
; Patent No. US20020111302A1  
; GENERAL INFORMATION:  
; APPLICANT: Tang, Y. Tom  
; APPLICANT: Zhou, Ping  
; APPLICANT: Goodrich, Ryle  
; APPLICANT: Liu, Chenghua  
; APPLICANT: Asundi, Vinod  
; APPLICANT: Wang, Jian-Rui  
; APPLICANT: Wang, Dunrui  
; APPLICANT: Yamazaki, Vicki  
; APPLICANT: Ujwal, Manusha L.  
; APPLICANT: Dmanac, Radoje T.  
; TITLE OF INVENTION: No. US20020111302A1e1 Nucleic Acids and  
; TITLE OF INVENTION: Polypeptides  
; FILE REFERENCE: 799  
; CURRENT APPLICATION NUMBER: US/09/728,952  
; CURRENT FILING DATE: 2000-11-30  
; NUMBER OF SEQ ID NOS: 101  
; SOFTWARE: pt\_FL\_genes Version 2.0  
; SEQ ID NO 28  
; LENGTH: 2753  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
; FEATURE:  
; NAME/KEY: CDS  
; LOCATION: (1)..(2418)  
US-09-728-952-28

Alignment Scores:  
Pred. No.: 0 Length: 2753  
Score: 380.00 Matches: 580  
Percent Similarity: 99.32% Conservative: 0  
Best Local Similarity: 99.32% Mismatches: 2  
Query Match: 51.63% Indels: 4  
Db: 9 Gaps: 0

US-10-045-815-4 (1-736) x US-09-728-952-28 (1-2753)

Qy 155 LysIleAsnLysLeuGluLysAlaValAlaAlaHisThrPheValGlyAsnPro 174  
Db 673 AAGATCACACAGTTGGAGAAAGCTGTGCTCGACACACACACTTCTTCGTGGGCAATCCT 732  
Qy 175 GluHisMetGluMetGlnGlnAsnLeuAspTyrTyrGlnThrMetSerGlyValLysGlu 194  
Db 733 GAGCACATGGAATGTCAGCAGAACCTAGACTATTACCAACCAATGCTCGGAGTGAAGGAG 792  
Qy 195 AlaAspPheLysAspLeuGlnThrGlnProHisMetGlnGluPheArgLeuGlyValArg 214  
Db 793 GCCGACTTCAGAGATCTTGAGACTCAACCCCATATGACAGAAATTCGACTGGGAGTGGCA 852  
Qy 215 LeuTyrSerGluGluGlnProGlnGluAlaValProHisLeuGluAlaLeuGlnGlu 234

Db 853 CTCCTACTCAGAGGAACACCCACAGAGCTGTGCCACCTAGAGCGCGCTGCAAGAA 912  
Qy 235 TyrPheValAlaTyrGluGluCysArgAlaIleLeuCysGluGlyProTyrAspTyrAspGly 254  
Db 913 TACTTTTGGGCTATGAGAGAGTGCCTCTCTCGAAGGGCCCTATGACTACATGCGC 972  
Qy 255 TyrAsnTyrLeuGluTyrAsnAlaAspLeuPheGlnAlaIleThrAspHisTyrIleGln 274  
Db 973 TACACTACTCTTGAGTCAACCGTGCACCTCTCCAGGCGCATCACAGATCATTTACATCCAG 1032  
Qy 275 ValLeuAsnCysLysGlnAsnCysValThrGlnLeuAlaSerHisProSerArgGluLys 294  
Db 1033 GTTCCTCAACTGTAAACAGAACTGTGTACGAGCTGTGCTCCACCAAGTCGAGAGAG 1092  
Qy 295 ProPheGluAspPheLeuProSerHisTyrAsnTyrLeuGlnPheAlaTyrTyrAsnIle 314  
Db 1093 CCCTTTGAGAGACTTCTCCCATCGCATTAATATATCTGCAGTTTGCCTACTATTAACATT 1152  
Qy 315 GlyAsnTyrThrGlnAlaGly-GluCysAlaLysThrTyrLeuLeuPhePheProAsnAs 334  
Db 1153 GGGAAATTATACACAGGCTGT-TGAAATGTGCCAAGACCTATCTCTCTTCTTCCCAATGA 1211  
Qy 334 pGluValMetAsnGlnAsnLeuAlaTyrTyrAlaAlaMetLeuGlyGluGluHisThrAr 354  
Db 1212 CGAGGTGATGAACCAAAATTTGGCCTATTATGCAGCTATGCTTGGAGAGAACACACACAG 1271  
Qy 354 gSerIleGlyProArgGluSerAlaLysGluTyrArgGlnArgSerLeuLeuGluLysGlu 374  
Db 1272 ATCCATCGGCCCCCGTGAGAGTGCCCAAGAGTACCGACAGCGAAGCTACTCGAAAAAGA 1331  
Qy 374 uLeuLeuPhePheAlaTyrAspValPheGlyIleProPheValAspProAspSerTrpTh 394  
Db 1332 ACTGCTTTTCTCGCTTATGATGTTTTTGGAAATCCCTTTTGGATTCGGATTCATGGAC 1391  
Qy 394 rProGluGluValIleProLysArgLeuGlnGluLysGlnLysSerGluArgGluThrAl 414  
Db 1392 TCAGNAGAGTGAATTCACAGAGATTCACAGAGAACACAGAGTCAGAACGGGAACAGC 1451  
Qy 414 aValArgIleSerGlnGluIleGlyAsnLeuMetLysGluIleGluThrLeuValGluGlu 434  
Db 1452 CGTACGCATCTCCAGGAGATTGGGAACCTTATGAAGGAAATTCGAGACCTTGTGGAAGA 1511  
Qy 434 uLysThrLysGluSerLeuAspValSerArgLeuThrArgGluGlyGlyProLeuLeuTy 454  
Db 1512 GAAGACCAAGAGTCACTGAGTGTGACGAGACTGACCGGAGAGGTGSCCCCTGCTGTA 1571  
Qy 454 rGluGlyIleSerLeuThrMetAsnSerLysLeuLeuAsnGlyTyr-GlnArgValValM 474  
Db 1572 TGAAGGCATCAGTCTCACCATGAATCCCAAACTCCTGAATGGTTTC-CCAGCGGGTGGTGA 1630  
Qy 474 etAspGlyValIleSerAspHisGluCysGlnGluLeuGlnArgLeuThrAsnValAlaA 494  
Db 1631 TGGACGGCGTAATCTCTGACCAAGAGTGTGAGAGCTGACAGACTGACCAATGTGGCAG 1690  
Qy 494 laThrSerGlyAspGlyTyrArgGlyGlnThrSerProHisThrProAsnGluLysPheT 514  
Db 1691 CAACCTCAGGAGATGCTACCGGGTTCAGACTCCCAACATCTCCCATGAAAGTTCT 1750  
Qy 514 YrGlyValThrValPheLysAlaLeuLysLeuGlyGlnGluLysValProLeuGlnS 534  
Db 1751 ATGGTGTCTACTGTCTTCAAAGCCCTCAAGCTCGGCAAGAGCAAGTTCCTCTGCAGA 1810  
Qy 534 erAlaHisLeuTyrTyrAsnValThrGluLysValArgArgIleMetGluSerTyrPheA 554  
Db 1811 GTGCCACCTGTACTATCAACGTCAGGAGAAAGTGGCGGCATCATGGAGTCTTCTTCC 1870  
Qy 554 rGluAspThrProLeuTyrPheSerTyrSerHisLeuValCysArgThrAlaIleGluG 574  
Db 1871 GCCTGATACGCCCTCTACTTTTCTTCTCATCTGCTGTCGCGACTGCGCATCGAAG 1930  
Qy 574 luValGlnAlaGluArgLysAspAspSerHisProValHisValAspAsnCysIleLeuA 594



Db 1931 AGGTCAGGACAGAGGAGGATGATGATCATCCAGTCCAGCGTGACAACTGCATCTCGA 1990  
QY 594 snAlaGluThrLeuValCysValLysGluProAlaTyrThrPheArgAspTyrSerA 614  
Db 1991 ATGCGGAGACCCCTCGTGTGTGTCAAAGAGACCCCGCCAGCCCTACACCTTCGCGACTACAGCG 2050  
QY 614 laileLeuTyrLeuAsnGlyAspPheAspGlyGlyAsnPheTyrPheThrGluLeuAspA 634  
Db 2051 CCATCTCTTACCTAATGGGACCTTCGATGGCGGAACTTTTATTTCTAGTGAATGGATG 2110  
QY 634 laLyThrValThrAlaGluValGlnProGlnCysGlyArgAlaValGlyPheSerSerg 654  
Db 2111 CCAAGACCGTGACGCGCAGAGGTGCAGCTCAGTGTGGAAGAGCGGTGGATTCTCTTCAG 2170  
QY 654 lyThrGluAsnProHisGlyValLysAlaValThrArgGlyGlnArgCysAlaIleAlaL 674  
Db 2171 GCACCTGAAGAACCCCATGGAGTGAAGGCTGTACCCAGGGGCGAGCGTGTGCCATCGCCC 2230  
QY 674 euTrpPheThrLeuAspProArgHisSerGluArgAspArgValGlnAlaAspAspLeuV 694  
Db 2231 TGTGGTTTCACTGACCTCGACACAGCGAGCGGACAGGGTGCAGGCGAGTACCTGG 2290  
QY 694 alLysMetLeuPheSerProGlnGluMetAspLeuSerGlnGluGlnProLeuAspAlag 714  
Db 2291 TGAAGATCTCTTACAGCCCGAGAGAGATGGACCTCTCCAGAGCAGCCCGCTGGATGCC 2350  
QY 714 lnGlnGlyProGluProAlaGlnGluSerLeuSerGlySerGluSerLysProLysA 734  
Db 2351 AGCAGGGCCCCCGAACCTGCACAAGAGTCTCTCTCAGGCGATGAATCGAAGCCCAAGG 2410  
QY 734 spGluLeu 736  
Db 2411 ATGAGCTA 2418

## RESULT 6

US-09-728-952-51  
; Sequence 51, Application US/09728952  
; Patent No. US2002011302A1  
; GENERAL INFORMATION:  
; APPLICANT: Tang, Y. Tom  
; APPLICANT: Zhou, Ping  
; APPLICANT: Goodrich, Ryle  
; APPLICANT: Liu, Chenghua  
; APPLICANT: Asundi, Vinod  
; APPLICANT: Wang, Jian-Rui  
; APPLICANT: Wang, Dunrui  
; APPLICANT: Yamazaki, Vicki  
; APPLICANT: Ujwal, Manusha L.  
; APPLICANT: Drmanac, Radoje T.  
; TITLE OF INVENTION: No. US2002011302A1el Nucleic Acids and  
; TITLE OF INVENTION: Polypeptides  
; FILE REFERENCE: 799  
; CURRENT APPLICATION NUMBER: US/09728,952  
; CURRENT FILING DATE: 2000-11-30  
; NUMBER OF SEQ ID NOS: 101  
; SOFTWARE: pt\_FL\_genes Version 2.0  
; SEQ ID NO 51  
; LENGTH: 2753  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
; FEATURE:  
; NAME/KEY: CDS  
; LOCATION: (1)...(2418)  
US-09-728-952-51

Alignment Scores:  
Pred. No.: 0  
Score: 380.00  
Percent Similarity: 99.32%  
Best Local Similarity: 99.32%  
Query Match: 51.63%  
DB: 9  
Length: 2753  
Matches: 580  
Conservative: 0  
Mismatch: 2  
Indels: 4  
Gaps: 0

US-10-045-815-4 (1-736) x US-09-728-952-51 (1-2753)

QY 155 LysIleAsnLysLeuGluLysAlaValAlaHisThrPhePheValGlyAsnPro 174  
Db 673 AGATCAACAGTTGGAGAAAGCTGTGTGTCGAGCACACACCTTCTTCTGTGGCAATCCT 732  
QY 175 GluHisMetGluMetGlnGlnAsnLeuAspTyrTyrGlnThrMetSerGlyValLysGlu 194  
Db 733 GAGCAGTGAATGAG 792  
QY 195 AlaAspPheLysAspLeuGluThrGlnProHisMetGlnGlnPheArgLeuGlyValArg 214  
Db 793 GCCGACTTCAAGGATCTTGAGACTCAACCCCATATGCAAGATTTTCGACTGGAGTGGGA 852  
QY 215 LeuTyrSerGluGluGlnProGlnGluAlaValProHisLeuGluAlaAlaLeuGlnGlu 234  
Db 853 CTCCTACTCAG 912  
QY 235 TyrPheValAlaTyrGluGluCysArgAlaLeuCysGluGlyProTyrAspTyrAspGly 254  
Db 913 TACTTTGTGGCCCTATGAGAGAGTGGCGTGGCCCTCTCGGAAGAGGCCCTATGACTACGATGGC 972  
QY 255 TyrAsnTyrLeuGluTyrAsnAlaAspLeuPheGlnAlaIleThrAspHisTyrIleGln 274  
Db 973 TACAACCTACTTGTAGTACAAAGCTGACCTCTTCCAGGCCATCACAGATCATTTACATCCAG 1032  
QY 275 ValLeuAsnCysLysGlnAsnCysValThrGluLeuAlaSerHisProSerArgGluLys 294  
Db 1033 GTCTCAACTGTAAAG 1092  
QY 295 PropheGluAspPheLeuProSerHisTyrAsnTyrLeuGlnPheAlaTyrTyrAsnIle 314  
Db 1093 CCCTTTGAAGACTTCTCCCATCGCATTAATATATCTGCAGTTTGCCTACTATAACATT 1152  
QY 315 GlyAsnTyrThrGlnAlaGly-GluCysAlaLysThrTyrLeuLeuPhePheProAsnAs 334  
Db 1153 GGGAAATTTATACACAGGCTGT-TGAATGTGCAAGAGAGAGAGAGAGAGAGAGAGAGAG 1211  
QY 334 pGluValMetAsnGlnAsnLeuAlaTyrTyrAlaAlaMetLeuGlyGluGluHisThrAr 354  
Db 1212 CGAGGTGATGAACCAAAATTTGGCCTATTATGACGATATGCTTGGAGAGAACACACAG 1271  
QY 354 gSerIleGlyProArgGluSerAlaLysGluTyrArgGlnArgSerLeuLeuGluLysG 374  
Db 1272 ATCCATCGGCCCCCGTGAGAGTGGCAAGAGTACCGACAGCGAAGAGAGAGAGAGAGAG 1331  
QY 374 uLeuLeuPhePheAlaTyrAspValPheGlyIleProPheValAspProAspSerTyrTh 394  
Db 1332 ACTGCTTTTCTCGCTTAATGATGTTTTTGGAAATTTCCCTTTGTGGATCCGATTCATGGAC 1391  
QY 394 rProGluValIleProLysArgLeuGlnGluLysGlnLysSerGluArgGluThrAl 414  
Db 1392 TCCAGAGAGAGTGAATTCCTCAAGAGATTCCAAAGAGAAACAGAAAGTCCAGAAACAGC 1451  
QY 414 aValArgIleSerGlnGluIleGlyAsnLeuMetLysGluIleGluThrLeuValGluG 434  
Db 1452 CGTACGATCTCCAGAGAGATTGGAACTTATGAAGAAATCGAGACCTTTGTGGAGAGAG 1511  
QY 434 uLysThrLysGluSerLeuAspValSerArgLeuThrArgGluGlyGlyProLeuLeuTy 454  
Db 1512 GAAGACCAAGAGTCACTGGATGTGAGCAGACTGACCCGGGAGAGTGGCCCCCTGCTGTA 1571  
QY 454 rGluGlyIleSerLeuThrMetAsnSerLysLeuLeuAsnGlyTyr-GlnArgValValM 474  
Db 1572 TGAAGGCGATCAGTCTCACCATGAATCAAACTCCTGAATGGTTC-CCAGCGGGTGGTGA 1630  
QY 474 etAspGlyValIleSerAspHisGluCysGlnGluLeuGlnArgLeuThrAsnValAlaA 494  
Db 1631 TGGACGGCGTAAATCTCTGACCACCGAGTGTGAGGAGCTGCAGAGAGTCAACCAATGTGGCAG 1690  
QY 494 laThrSerGlyAspGlyTyrArgGlyGlnThrSerProHisThrProAsnGluLysPheT 514  
Db 1691 CAACCTCAGGAGATGGCTACCGGGTTCAGACCTCCCAACATCTCCCAATGAAAGATTCT 1750



Qy 514 YrGlyValThrValPheLysAlaLeuLysLeuGlyGlnGluGlyLysValProLeuGlnS 534  
Db 1751 ATGGTGTCACTGTCTTCAAGAGCCCTCAAGCTGGGCAAGAGCAAGATTCCTCTCGAGA 1810  
Qy 534 exAlaHisLeuTyrTyrAsnValThrGluLysValArgArgilleMetGluSerTyrPheA 554  
Db 1811 GTGCCACCTGTACTACAACTGACGAGAGAGTGGCGGCATCATGGAGTCTACTTCC 1870  
Qy 554 rgLeuAspThrProLeuTyrPheSerTyrSerHisLeuValCysArgThrAlaIleGluG 574  
Db 1871 GCCTGGATAGCCGCTCTACTTTCTCTACTCTCATCTGTGTGTCGGCAGCTGCCATCGAAG 1930  
Qy 574 luValGlnAlaGluuLysAspAspSerHisProValHisValAspAsnCysIleLeuA 594  
Db 1931 AGGTCCAGGACAGAGAGAGATGATGATCTCATCGATCCAGCTGGACAACTGGCATCTGA 1990  
Qy 594 snAlaGluThrLeuValCysValLysGluProProAlaTyrThrPheArgAspTyrSerA 614  
Db 1991 ATGCCGAGACCTCGTGTGTGTCAAAGAGAGCCCGCAGCCTACACCTTCGGCGACTACAGCG 2050  
Qy 614 lalleLeuTyrLeuAsnGlyAspPheAspGlyGlyAsnPheTyrPheThrGluLeuAspA 634  
Db 2051 CCATCTCTTAACTAAATGGGAGCTTCGATGGCGGAAACTTTTATTTCACTGAACCTGGATG 2110  
Qy 634 laLysThrValThrAlaGluValGlnProGlnCysGlyArgAlaValGlyPheSerSerG 654  
Db 2111 CCAGACCGTGACGCGAGAGGTGAGCTCAGTGTGGAAGAGCGGTGGGATTTCTTTTCAG 2170  
Qy 654 lyThrGluAsnProHisGlyValLysAlaValThrArgGlyGlnArgCysAlaIleAlaL 674  
Db 2171 GCACTGAAACCCACATGGAGTGAAGCTGTCCACAGCGGCGACGCTGTGCCATCGCCC 2230  
Qy 674 euTTPheThrLeuAspProArgHisSerGluArgAspArgValGlnAlaAspAspLeuV 694  
Db 2231 TGTGGTTCACCTGGACCTCGACACAGCGAGCGGACAGGTGTCAGGACGATGACCTGG 2290  
Qy 694 allysMetLeuPheSerProGluGluMetAspLeuSerGlnGluGlnProLeuAspAlaG 714  
Db 2291 TGAAGATCTCTTCAGCCACAGAGATGGACCTCTCCAGGAGCAGCCCTGGATGCC 2350  
Qy 714 lngInGlyProProGluProAlaGlnGlnSerLeuSerGlySerGluSerLysProLysA 734  
Db 2351 AGCAGGGCCCCCGAACCTGCACAAGAGTCTCTCTCAGGCAGTGAATCGAAGCCCCAAGG 2410  
Qy 734 spGluLeu 736  
Db 2411 ATGAGCTA 2418

RESULT 7

US-10-045-815-1  
; Sequence 1, Application US/10045815  
; Publication No. US20020160498A1  
; GENERAL INFORMATION:  
; APPLICANT: Wadhwa, Renu  
; APPLICANT: Sugihara, Takashi  
; TITLE OF INVENTION: TUMOR SUPPRESSOR GENE  
; FILE REFERENCE: 06501-091001  
; CURRENT APPLICATION NUMBER: US/10/045,815  
; CURRENT FILING DATE: 2001-10-26  
; PRIOR APPLICATION NUMBER: PCT/JPO00/02731  
; PRIOR FILING DATE: 2000-04-26  
; PRIOR APPLICATION NUMBER: JP 11/118806  
; PRIOR FILING DATE: 1999-04-26  
; NUMBER OF SEQ ID NOS: 15  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 1  
; LENGTH: 2829  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
; FEATURE:  
; NAME/KEY: CDS

i LOCATION: (52)...(1140)

US-10-045-815-1

Alignment Scores:  
Pred. No.: 0 Length: 2829  
Score: 360.00 Matches: 360  
Percent Similarity: 100.00% Conservative: 0  
Best Local Similarity: 100.00% Mismatches: 0  
Query Match: 48.91% Indels: 0  
DB: 14 Gaps: 0

US-10-045-815-4 (1-736) x US-10-045-815-1 (1-2829)

Qy 1 MetAlaValArgAlaLeuLysLeuLeuThrThrLeuLeuAlaValValAlaAlaSer 20  
Db 52 ATGGCGGTACGCGCGTGAAGCTGTGACCACTGCTGGTGTGGTGGCGCTCCCTCC 111  
Qy 21 GlnAlaGluValGluSerGluAlaGlyTTPGlyMetValThrProAspLeuLeuPheAla 40  
Db 112 CAAGCGAGGTGCGATTCGAGGAGGATGGGCGATGGTGACGCTGATCTGCTCTCGCC 171  
Qy 41 GluGlyThrAlaAlaTyrAlaArgGlyAspTTPProGlyValValLeuSerMetGluArg 60  
Db 172 GAGGGGACCGCAGCTACGCGCGGGGACTGGCCCGGGGTGGTCTCTGAGCATGGAACGG 231  
Qy 61 AlaLeuArgSerArgAlaAlaLeuArgAlaLeuArgLeuArgCysArgThrGlnCysAla 80  
Db 232 GCGCTGCGCTCCGCGGACGCTTCGCGCCCTTCGCTGCGCGCACCCAGGTGTGCC 291  
Qy 81 AlaAspPheProTTPGluLeuAspProAspTTPSerProSerProAlaGlnAlaSerGly 100  
Db 292 GCGCATTTCCGTGGAGGTGGACCCGACCTGGTCCCCCAGCCGCGCCGAGGCTCGGCG 351  
Qy 101 AlaGlyAlaLeuArgAspLeuSerPhePheGlyGlyLeuLeuArgArgAlaAlaCysLeu 120  
Db 352 GCGCGCGCCTGCGCGACCTGAGCTTCTTCGGGGGCTTCTGCTGCGCTGCTGCTGCTG 411  
Qy 121 ArgArgCysLeuGlyProProAlaAlaHisSerLeuSerGluGluMetGluLeuGluPhe 140  
Db 412 CGCGCTGCTGCGCGCGCGCGCGCCCTGCTGCGCGAGAGATGGAGTGGAGTTTC 471  
Qy 141 ArgLysArgSerProTyrAsnTyrLeuGlnValAlaTyrPheLysIleAsnLysLeuGlu 160  
Db 472 CGCAAGCGGAGCGCCCTACAACTACCTGAGGTGGCTACTTCAAGATCAACAAGTTGGAG 531  
Qy 161 LysAlaValAlaAlaHisThrPhePheValGlyAsnProGluHisMetGluMetGln 180  
Db 532 AAAGCTGTTGCTGCGACACACACCTTCTTCGTGGGCAATCTTGAGCACATGGAAATGCCAG 591  
Qy 181 GlnAsnLeuAspTyrTyrGlnThrMetSerGlyValLysGluAlaAspPheLysAspLeu 200  
Db 592 CAGAACTAGACTATTTCNAACCACTGTCTGGAGTGAAGAGCGCCGACTTCAAGGATCTT 651  
Qy 201 GluThrGlnProHisMetGlnGluPheArgLeuGlyValArgLeuTyrSerGluGluGln 220  
Db 652 GAGACTCAACCCCATATGCAAGAAATTTGACTGGAGTGGACTCTACTCAGAGGAACAG 711  
Qy 221 ProGlnGluAlaValProHisLeuGluAlaAlaLeuGlnGluTyrPheValAlaTyrGlu 240  
Db 712 CCACAGGAAGCTGTGCCCCACCTAGAGGCGCGCTGCAAGATACCTTTGTGGCTATGAG 771  
Qy 241 GluCysArgAlaLeuCysGluGlyProTyrAspTyrAspGlyTyrAsnTyrLeuGluTyr 260  
Db 772 GAGTGGCTGCTCTCTGCGAAGGCGCCCTATGACTAGCTGGCTACAACCTACCTTGAATAC 831  
Qy 261 AsnAlaAspLeuPheGlnAlaIleThrAspHisTyrIleGlnValLeuAsnCysLysGln 280  
Db 832 AACGCTACCTCTTCCAGGCGCATCACAGATCATTATCATCCAGGTCTCTCACTGTAGCAG 891  
Qy 281 AsnCysValThrGluLeuAlaSerHisProSerArgGluLysProPheGluAspPheLeu 300  
Db 892 AACTGTGTACGGAGCTTGTCTTCCACCACCAAGTCGAGAGAGCCCTTTTGAAGACTTCCCTC 951

QY 301 ProSerHisTyrAsnTyrLeuGlnPheAlaTyrTyrAsnIleGlyAsnTyrThrGlnAla 320  
Db 952 CCATCCCATTAATTAATTAATTCGAGTTTGCTACTATACATTTGGGAATTAACAGAGCT 1011  
QY 321 GlyGluCysAlaIysThrTyrLeuLeuPhePheProAsnAspGluValMetAsnGlnAsn 340  
Db 1012 GGTGAATGTGCCAAGACCTATCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 1071  
QY 341 LeuAlaTyrTyrAlaAlaMetLeuGlyGluGluHisThrArgSerIleGlyProArgGlu 360  
Db 1072 TTGGCCATTAATGACGCTATGCTTGGAGAAGAACACACACAGATCCATCGGCCCGCTGAG 1131

## RESULT 8

US-10-257-174-1  
; Sequence 1, Application US/10257174  
; Publication No. US20040034194A1  
; GENERAL INFORMATION:  
; APPLICANT: Agarwal, Pankaj  
; APPLICANT: Murdock, Paul R.  
; APPLICANT: Rizvi, Safia K.  
; APPLICANT: Smith, Randall F.  
; APPLICANT: Xiang, Zhaoqing  
; TITLE OF INVENTION: NOVEL COMPOUNDS  
; FILE REFERENCE: GP50022  
; CURRENT APPLICATION NUMBER: US/10/257,174  
; CURRENT FILING DATE: 2002-10-10  
; PRIOR APPLICATION NUMBER: PCT/US01/11797  
; PRIOR FILING DATE: 2001-04-11  
; PRIOR APPLICATION NUMBER: 60/196,603  
; PRIOR FILING DATE: 2000-04-13  
; PRIOR APPLICATION NUMBER: 60/199,417  
; PRIOR FILING DATE: 2000-04-24  
; NUMBER OF SEQ ID NOS: 48  
; SOFTWARE: FastSeq for Windows Version 3.0  
; SEQ ID NO 1  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
US-10-257-174-1

## Alignment Scores:

Pred. No.: 5,156-274 Length: 2127  
Score: 275.00 Matches: 375  
Percent Similarity: 99.47% Conservative: 0  
Best Local Similarity: 99.47% Mismatches: 1  
Query Match: 37.36% Indels: 2  
DB: 13 Gaps: 0

US-10-045-815-4 (1-736) x US-10-257-174-1 (1-2127)

QY 361 SerAlaIysGluTyrArgGlnArgSerLeuLeuGlyGluLeuLeuPheAlaTyr 380  
Db 997 AGTGCCAAAGAGTACCGACGGAAGGCTACTGGAAAAAGAACTGCTTTCTTCGCTTAT 1056  
QY 381 AspValPheGlyIleProPheValAspProAspSerTyrThrProGluGluValIlePro 400  
Db 1057 GATGTTTTTGAATTCCTTTTGTCATCCGATTCATGAGTTCAGAGTCCAGAAAGTGTATCCC 1116  
QY 401 LysArgLeuGlnGluLysGlnLysSerGluArgGluThrAlaValArgIleSerGlnGlu 420  
Db 1117 AAGAGATTGCAAGAGAAACAGAACTCAGAACGGGAAACAGCCGTACCATCTCCACAGGAG 1176  
QY 421 IleGlyAsnLeuMetLysGluIleGluThrLeuValGluGluLysThrLysGluSerLeu 440  
Db 1177 ATTGGGAACCTTATGAAGGAATCGACCTTGTGGAAGAGAACCAAGGAGTCACTG 1236  
QY 441 AspValSerArgLeuThrArgGluGlyGlyProLeuLeuTyrGluGlyIleSerLeuThr 460  
Db 1237 GATGTGACGAGACTGACCCGGAAGGTGGGCCCTGCTGTATGAAGGCATCAGTCTCACC 1296  
QY 461 MetAsnSerLysLeuLeuAsnGlyTyr-GlnArgValValMetAspGlyValIleSerAs 480  
Db 1297 ATGAACCTCCAACTCTCTGAATGGTTC-CCAGCGGGGTGGTGTGACGGCGGTATCTCTGA 1355

QY 480 pHisGluCysGlnGluLeuGlnArgLeuThrAsnValAlaAlaThrSerGlyAspGlyTy 500  
Db 1356 CCACGAGTGTCCAGAGCTGCAGAGACTGCACCAATGTGGCAGCAACCTCAGGAGATGGCTA 1415  
QY 500 rAtsGlyGlnThrSerProHisThrProGlnGluLysPheTyrGlyValThrValPheTy 520  
Db 1416 CCGGGGTACAGACTCCCATCACTACTCCCAATGAAAAGTTCTATGGTGTCTCAGTCTCTCAA 1475  
QY 520 sAlaLeuLysLeuGlyGlnGluLysValProLeuGlnSerAlaHisLeuTyrTyrAs 540  
Db 1476 AGCCCTCAAGCTGGGCAAGAGGCAAGTTCTCTTCGAGAGTGCCCACTGTACTACAA 1535  
QY 540 nValThrGluLysValArgArgIleMetGluSerTyrPheArgLeuAspThrProLeuTy 560  
Db 1536 CGTGACGGAGAGGTGCGGCGCATCATGGAGTCTTACTTCCGCTGGATACGCCCTCTTA 1595  
QY 560 rPheSerTyrSerHisLeuValCysArgThrAlaIleGluGluValGlnAlaGluArgTy 580  
Db 1596 CTTTTCTACTCTCACTCTGGTGTGGCGCACTGCCATCGAAGAGGTCCAGGACAGAGAA 1655  
QY 580 sAspAspSerHisProValHisValAspAsnCysIleLeuAsnAlaGluThrLeuValCy 600  
Db 1656 GGATGATAGTATCCAGTCCAGTCCGACCACTGCATCTTGAATGCCGAGACCTCGTGTG 1715  
QY 600 sValLysGluProProAlaTyrThrPheArgAspTyrSerAlaIleLeuTyrLeuAsnGl 620  
Db 1716 TGTCAAAGAGCCCCAGCCCTACACCTTCGCGCACTACAGCGCCTCTTACTTAATGG 1775  
QY 620 yAspPheAspGlyGlyAsnPheTyrPheThrGluLeuAspAlaIysThrValThrAlaGl 640  
Db 1776 GGACTTCGATGGCGAAACTTTTATTTCACTGAATGGATGCAAGACCGTGCAGCGAGA 1835  
QY 640 uValGlnProGlnCysGlyArgAlaValGlyPheSerGlyThrGluAsnProHisGl 660  
Db 1836 GGTGCAGCCTCAGTGTGGAAGAGCGTGGGATCTCTTCAGGCATGAAAAACCCATGG 1895  
QY 660 yValLysAlaValThrArgGlyGlnArgCysAlaIleAlaLeuTyrPheThrLeuAspPr 680  
Db 1896 AGTGAAGGCTGTACCCAGGGGCGCAGCGCTGTGCCATCGCCCTGTGTTCACTCCAGACC 1955  
QY 680 oArgHisSerGluArgAspArgValGlnAlaAspAspLeuValLysMetLeuPheSerPr 700  
Db 1956 TCGACACAGCAGCGGGACAGGGTGCAGCAGATACCTGTGTGAAGATGCTCTTTCAGGCC 2015  
QY 700 oGluGluMetAspLeuSerGlnGluProLeuAspAlaGlnGlnGlyProProGluPr 720  
Db 2016 AGAAGAGATGACCTCTCCAGGACAGCCCTCGATGCCAGCAGGGTCCCCCGGAACC 2075  
QY 720 oAlaGlnGluSerLeuSerGlySerGluSerLysProLysAspGluLeu 736  
Db 2076 TGCACAGAGTCTCTCTCAGGCAGTGAATCGAAGCCCAAGGATGAGCTA 2124

## RESULT 9

US-10-302-172-227  
; Sequence 227, Application US/10302172  
; Publication No. US20040053250A1  
; GENERAL INFORMATION:  
; APPLICANT: Tang, Y. Tom  
; APPLICANT: Xue, Aidong J.  
; APPLICANT: Drmanac, Radolje T.  
; TITLE OF INVENTION: No. US20040053250A1el Arginine-rich Protein-like Nucleic Acids and  
; FILE REFERENCE: 803\_1NCP  
; CURRENT APPLICATION NUMBER: US/10/302,172  
; CURRENT FILING DATE: 2002-11-21  
; PRIOR APPLICATION NUMBER: US 10/225,251  
; PRIOR FILING DATE: 2002-08-20  
; PRIOR APPLICATION NUMBER: PCT US02/05095  
; PRIOR FILING DATE: 2002-03-05  
; PRIOR APPLICATION NUMBER: US 09/799,451  
; PRIOR FILING DATE: 2001-03-05  
; NUMBER OF SEQ ID NOS: 950

```
; SOFTWARE: pt_FL_genes Version 2.0
; SEQ ID NO 227
; LENGTH: 2152
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (32)..(2149)
US-10-302-172-227

Alignment Scores:
Pred. No.: 6,14e-242 Length: 2152
Score: 244.00 Matches: 358
Percent Similarity: 98.90% Conservative: 0
Best Local Similarity: 98.90% Mismatches: 2
Indels: 4
Query Match: 33.15%
DB: 13 Gaps: 0

US-10-045-815-4 (1-736) x US-10-302-172-227 (1-2152)

Qy 1 MetAlaValArgAlaLeuLysLeuThrThrLeuLeuAlaValValAlaAlaSer 20
Db 32 ATGGCGGTACGGCGTTGAAGTGTGACACACATGCTGGCTGCGGTGGCGCTGGCTCC 91
Qy 21 GlnAlaGluValGluSerGluAlaGlyTyrGlyMetValThrProAspLeuPheAla 40
Db 92 CAAGCCAGGTTCAGTCCGAGGAGGATGGGATGGTGGCTGATCTGCTCTTCGCC 151
Qy 41 GluGlyThrAlaAlaTyrAlaArgGlyAspTyrProGlyValValLeuSerMetGluArg 60
Db 152 GAGGGAGCCGAGCTACGGCGCGGGGACTGGCGCGGGTGGTCTTGGCTGATCTTCGCC 211
Qy 61 AlaLeuArgSerArgAlaAlaLeuArgAlaLeuArgLeuArgCysArgThrGlnCysAla 80
Db 212 GCGTGGCGCTCCCGGGGAGCCCTCCGGCCCTTCGCTGGCTGGCTGGCTGGCTGGCT 271
Qy 81 AlaAspPheProTyrGluLeuAspProAspTyrSerProSerProAlaGlnAlaSerGly 100
Db 272 GCCGACTTCGCGTGGAGTGGACCCCGACTGGTCCCGCCAGCGCGGCGCTGGCG 331
Qy 101 AlaGlyAlaLeuArgAspLeuSerPheGlyGlyLeuLeuArgAlaAlaCysLe 120
Db 332 GCCGC-CCGCCCTGGCGACCTGAGCTTCTCGGGGGGCTTCTGGCTGGCTGGCTGGCT 390
Qy 120 uArgArgCysLeuGlyProProAlaAlaHisSerLeuSerGluLeuMetGluLeuGluPh 140
Db 391 CGCGCGTGGCTGGCGCGCGCGCGCCACTCGCTCAGCGAGAGATGGAGTGGAGTT 450
Qy 140 eArgLysArgSerProTyrAsnTyrLeuGlnValAlaTyrPheLysLeuLysLeuG 160
Db 451 CCGCAAGCGGAGCCCTACAACTACCTGAGGTGGCTTCTTCAAGATCAACAAGTTGGA 510
Qy 160 uLysAlaValAlaAlaAlaHisThrPhePheValGlyValAsnProGluHisMetGluMetG 180
Db 511 GAAAGCTGTGCTGGAGCACACACCTTCTTCTGGGCAATCTGAGCACATGGAAATGCA 570
Qy 180 nGlnAsnLeuAspTyrTyrGlnThrMetSerGlyValLysGluAlaAspPheLysAspLe 200
Db 571 GCAGACCTAGACTATTACCAAAACCATCTCTGGAGTGAAGGAGCGCGACTTCAAGGATCT 630
Qy 200 uGluThrGlnProHisMetGlnGluPheArgLeuGlyValArgLeuTyrSerGluGlu 220
Db 631 TGAGACTCAACCCCATATGCAAGAAATTTGAGATGGAGTGGACTCTACTCAGAGGAACA 690
Qy 220 nProGlnGluAlaValProHisLeuGluAlaAlaLeuGlnGluTyrPheValAlaTyrG 240
Db 691 GCCACAGAGNAGCTGTGCCCACTAGAGCGCGCTGCAAGAATACTTTTGGCTATGA 750
Qy 240 uGluCysArgAlaLeuCysGluGlyProTyrAspTyrAspGlyTyrAsnTyrLeuGluT 260
Db 751 GGAGTGGCGTGGCTCTGCAAGGGCCCTATGACTACGATGGCTACAACTACCTTGGATA 810
Qy 260 rAsnAlaAspLeuPheGlnAlaIleThrAspHisTyrIleGlnValLeuAsnCysLysG 280
```

```
Db 811 CAAGCTGACCTCTTCCAGGCCATCAGACATCAATACATCCAGTCTCAACTGTAAGCA 870
Qy 280 rAsnCysValThrGluLeuAlaSerHisProSerArgGluLysProPheGluAspPheLe 300
Db 871 GAACCTGTGTACGGAGCTTGTCTCCACCCCAAGTCGAGAGAAGCCCTTTGAAGACTTCC 930
Qy 300 uProSerHisTyrAsnTyrLeuGlnPheAlaTyrTyrAsnIleGlyAsnTyrThrGlnAl 320
Db 931 CCCATCGCATTAATATTAATCTGCGTTCGCTACTATAACATTGGGAATTATACACAGC 990
Qy 320 aGly-GluCysAlaLysThrTyrLeuLeuPhePheProAsnAspGluValMetAsnGlnA 340
Db 991 TGT-TGAATGTGCCAAGACCTATCTTCTCTTCCCAATGACAGGTGATGACACCAA 1049
Qy 340 snLeuAlaTyrTyrAlaAlaMetLeuGlyGluGluHisThrArgSerIleGlyProArg 360
Db 1050 ATTTGGCTTATTATGCACTATGCTTTGGAGAAGAACAACACAGATCCATCGGCCCTG 1109
Qy 360 lu 360
Db 1110 AG 1111

RESULT 10
US-10-302-172-226
; Sequence 226, Application US/10302172
; Publication No. US20040053250A1
; GENERAL INFORMATION:
; APPLICANT: Tang, Y. Tom
; APPLICANT: Xue, Aidong J.
; APPLICANT: Drmanac, Radoje T.
; TITLE OF INVENTION: No. US20040053250A1el Arginine-rich Protein-like Nucleic Acids and
; TITLE OF INVENTION: Polypeptides
; FILE REFERENCE: 803 1CNCp
; CURRENT APPLICATION NUMBER: US/10/302,172
; CURRENT FILING DATE: 2002-11-21
; PRIOR APPLICATION NUMBER: US 10/225,251
; PRIOR FILING DATE: 2002-08-20
; PRIOR APPLICATION NUMBER: PCT US02/05095
; PRIOR FILING DATE: 2002-03-05
; PRIOR APPLICATION NUMBER: US 09/799,451
; PRIOR FILING DATE: 2001-03-05
; NUMBER OF SEQ ID NOS: 950
; SOFTWARE: pt_FL_genes Version 2.0
; SEQ ID NO 226
; LENGTH: 1437
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (32)..(1255)
US-10-302-172-226

Alignment Scores:
Pred. No.: 3,13e-216 Length: 1437
Score: 219.00 Matches: 406
Percent Similarity: 99.02% Conservative: 0
Best Local Similarity: 99.02% Mismatches: 2
Query Match: 29.76%
DB: 13 Gaps: 0

US-10-045-815-4 (1-736) x US-10-302-172-226 (1-1437)

Qy 1 MetAlaValArgAlaLeuLysLeuThrThrLeuLeuAlaValValAlaAlaSer 20
Db 32 ATGGCGGTACGGCGTTGAAGTGTGACACACATGCTGGCTGCGGTGGCGCTGGCTCC 91
Qy 21 GlnAlaGluValGluSerGluAlaGlyTyrGlyMetValThrProAspLeuPheAla 40
Db 92 CAAGCCAGGTTCAGTCCGAGGAGGATGGGATGGTGGCTGATCTGCTCTTCGCC 151
Qy 41 GluGlyThrAlaAlaTyrAlaArgGlyAspTyrProGlyValValLeuSerMetGluArg 60
```

Db	152	GAGGGGACCGCAGCCTACGCGCGCGGGGACTCGCCCGGGGTGGTCTGTAGCATGAAACGG	211
Qy	61	AlaLeuArgSerArgAlaAlaLeuArgAlaLeuArgLeuArgCysArgThrGlnCysAla	80
Db	212	GGCTGGCGTCCCGGGCAGCCCTCCGCGCCCTTCGCTGGCTGGCGCACCCAGTGTGCC	271
Qy	81	AlaAspPheProTrrPgluLeuAspProAspTrrPserProSerProAlaGlnAlaSerGly	100
Db	272	GCCGACTTCCCGTGGAGCTGGACCCGACTGGTCCCGCAGCCGCGCCAGGCTTCGGGC	331
Qy	101	AlaGly-AlaLeuArgAspLeuSerPhePheGlyGlyLeuLeuArgArgAlaAlaCysLe	120
Db	332	GCCGC-CGCCCTGCCGACCTGAGCTTCTTTCGGGGCCCTTCGCGTGGCGTGGCTGCT	390
Qy	120	uArgArgCysLeuGlyProProAlaAlaHisSerLeuSerGluGluMetGluLeuGluPh	140
Db	391	GCGCGCTGCCTCGCGCGCGCGCGCCGCTCGCTCAGCGAAAGAGATGGAGCTGGAGTT	450
Qy	140	eArgLysArgSerProTyrAsnTyrLeuGluValAlaTyrPheLysIleAsnLysLeuGl	160
Db	451	CCGAAAGCGGAGCCCTCAACTACTCGAGCTCGCTTCTTCAAGATCAACAAGTTCGA	510
Qy	160	uLysAlaValAlaAlaAlaHisThrPhePheValGlyAsnProGluHisMetGluMetGl	180
Db	511	GAAAGCTGTGTCGAGCACACACCTCTCTCGTGGCAATCCTGAGCACATGGAAATGCA	570
Qy	180	nGlnAsnLeuAspTyrTyrGlnThrMetSerGlyValLysGluAlaAspPheLysAspLe	200
Db	571	GCAGAACTAGACTATTACCAAAACCATGCTCTGGAGTGAAGAGGCGCGACTTCAAGGAT	630
Qy	200	uGluThrGlnProHisMetGlnGluPheArgLeuGlyValArgLeuTyrSerGluGluGl	220
Db	631	TGAGACTCAACCCCATATGCAAGAAATTCGACTGGAGTGGCATCTACTCTGAGAGAA	690
Qy	220	nProGlnGluAlaValProHisLeuGluAlaAlaLeuGlnGluTyrPheValAlaTyrGl	240
Db	691	GCCACAGAAAGCTGTGCCCAACCTTAGAGGCGCGCTGCCAAGATACTTTGTGGCGTATGA	750
Qy	240	uGluCysArgAlaLeuCysGluGlyProTyrAspTyrAspGlyTyrAsnTyrLeuGluTy	260
Db	751	GGAGTGGCGTCCCTCTGGAGAGGGCCCTATGACTAGATGGCTACAACTACCTTTGATGA	810
Qy	260	rAsnAlaAspLeuPheGlnAlaIleThrAspHisTyrIleGlnValLeuAsnCysLysGl	280
Db	811	CAACGCTGAGCTCTTCCAGGCGCATCAGACTCATTAATCATCAGGTCTCAACTGTAAGCA	870
Qy	280	nAsnCysValThrGluLeuAlaSerHisProSerArgGluLysProPheGluAspPheLe	300
Db	871	GAACCTGTGCAGGAGCTTGCTTCCCAACCAAGTCGAGAGAGCCCTTTGAAAGACTTCCT	930
Qy	300	uProSerHisTyrAsnTyrLeuGlnPheAlaTyrTyrAsnIleGlyAsnTyrThrGlnAl	320
Db	931	CCCATCGCATTAATAATTATCTGCAGTTTGCTACTATAACATTGGGAATTTATACAGAGC	990
Qy	320	aGly-GluCysAlaLysThrTyrLeuLeuPhePheProAsnAspGluValMetAsnGlnA	340
Db	991	TGT-TGAATGTGCCAAGACCTATCTTCTCTTCTTCCCAATGACGAGGTGATGAACCAAA	1049
Qy	340	snLeuAlaTyrTyrAlaAlaMetLeuGlyGluGluHisThrArgSerIleGlyProArgG	360
Db	1050	ATTGTGCTATTATGAGCTATGCTTGGAGAAGAACACACACAGATCCATCGGCCCCCGTG	1109
Qy	360	luSerAlaLysGluTyrArgGlnArgSerLeuLeuGluLysGluLeuLeuPhePheAlaI	380
Db	1110	AGAGTCCCAAGGAGTACCGACAGAGCCTATCGGAAAAAAGAACTGCTTTTCTTGGCTT	1169
Qy	380	YrAspValPheGlyIleProPheValAspProAspSerTrpThrProGluGluValIleP	400
Db	1170	ATGATGTTTGAATTCCTTTGTGGATCCGATTCATGGACTCCAGAAAGAGTGAATTC	1229
Qy	400	rolLysArgLeuGlnLysGlnLys	408
Db	1230	CCAAGAGATTGCAAGAGAAACAGAG	1255

RESULT 11

```

US-10-085-783A-6419
; Sequence 6419, Application US/10085783A
; Publication No. US20040037841A1
; GENERAL INFORMATION:
; APPLICANT: ChondroGene Inc.
; APPLICANT: Liew, C.C.
; TITLE OF INVENTION: Compositions and Methods Relating to Osteoarthritis
; FILE REFERENCE: 4231/2002
; CURRENT APPLICATION NUMBER: US/10/085,783A
; CURRENT FILING DATE: 2002-02-28
; PRIOR APPLICATION NUMBER: US 60/305,340
; PRIOR FILING DATE: 2001-07-13
; PRIOR APPLICATION NUMBER: US 60/275,017
; PRIOR FILING DATE: 2001-03-12
; PRIOR APPLICATION NUMBER: US 60/271,955
; PRIOR FILING DATE: 2001-02-28
; NUMBER OF SEQ ID NOS: 58994
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 6419
; LENGTH: 337
; TYPE: DNA
; ORGANISM: Human
US-10-085-783A-6419

Alignment Scores:
Pred. No.: 4,61e-105 Length: 337
Score: 111.00 Matches: 111
Percent Similarity: 100.00% Conservatives: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 15.08% Indels: 0
DB: 13 Gaps: 0

```

```

; PRIORITY: 1
; NAME/KEY: misc_feature
; LOCATION: 2282
; OTHER INFORMATION: n = A,T,C or G
US-10-045-815-7

Alignment Scores:
Pred. No.: 1.29e-83
Score: 91.00
Percent Similarity: 100.00%
Best Local Similarity: 100.00%
Query Match: 12.36%
DB: 14

US-10-045-815-4 (1-736) x US-10-045-815-7 (1-2322)
QY 612 TyrSerAlaIleLeuTyrLeuAsnGlyAspPheAspClyVlyAsnPhetYrPheThrGlu 631
DB 1757 TACAGCGCATCTTTTACCTCAATGGCGACTTTCATGGAGAACTTTTACTTTCACAGAA 1816
QY 632 LeuAspAlaLysThrValThrAlaGluValGlnProGlnCysGlyArgAlaValGlyPhe 651
DB 1817 CTAGATGCCAAGACTGTGACGGCAGAGGTGCAGCCCACTGTGGAGGGGTGTGGGATT 1876
QY 652 SerSerGlyThrGluAsnProHisGlyValLysAlaValThrArgGlyGlnArgCysAla 671
DB 1877 TCTTCTGGCACTGAGAAACCCACATGGAGTGAAGCTGTCCACAGGGGGCAGCGCTGCGCC 1936
QY 672 IleAlaLeuTyrPheThrLeuAspProArgHisSerGluArgAspArgValGlnAlaAsp 691
DB 1937 ATCGCCCTGTGGTTCACGCTGGATCCTCGGCACAGTGCAGAGACAGAGAGAGAGAT 1996
QY 692 AspLeuValLysMetLeuPheSerProGluGlu 702
DB 1997 GACCTGTGTGAAGATGCTGTTTCAGCCCAAGAGAG 2029

RESULT 14
US-10-045-815-5
; Sequence 5, Application US/10045815
; Publication NO. US20020160498A1
; GENERAL INFORMATION:
; APPLICANT: Wadhwah, Renu
; APPLICANT: Sugihara, Takashi
; APPLICANT: Ohide, Akiko
; TITLE OF INVENTION: TUMOR SUPPRESSOR GENE
; FILE REFERENCE: 06501-091001
; CURRENT APPLICATION NUMBER: US/10/045,815
; CURRENT FILING DATE: 2001-10-26
; PRIOR APPLICATION NUMBER: PCT/JP00/02731
; PRIOR FILING DATE: 2000-04-26
; PRIOR APPLICATION NUMBER: JP 11/118806
; PRIOR FILING DATE: 1999-04-26
; NUMBER OF SEQ ID NOS: 15
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 5
; LENGTH: 2416
; TYPE: DNA
; ORGANISM: Mus musculus
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (12)...(2252)
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 2376_
; OTHER INFORMATION: n = A,T,C or G
US-10-045-815-5

Alignment Scores:
Pred. No.: 1.33e-83
Score: 91.00
Percent Similarity: 100.00%
Best Local Similarity: 100.00%
Query Match: 12.36%
DB: 14

; PRIORITY: 1
; NAME/KEY: misc_feature
; LOCATION: 2282
; OTHER INFORMATION: n = A,T,C or G
US-10-045-815-7

Alignment Scores:
Pred. No.: 1.29e-83
Score: 91.00
Percent Similarity: 100.00%
Best Local Similarity: 100.00%
Query Match: 12.36%
DB: 14

US-10-045-815-4 (1-736) x US-10-045-815-7 (1-2322)
QY 612 TyrSerAlaIleLeuTyrLeuAsnGlyAspPheAspClyVlyAsnPhetYrPheThrGlu 631
DB 1757 TACAGCGCATCTTTTACCTCAATGGCGACTTTCATGGAGAACTTTTACTTTCACAGAA 1816
QY 632 LeuAspAlaLysThrValThrAlaGluValGlnProGlnCysGlyArgAlaValGlyPhe 651
DB 1817 CTAGATGCCAAGACTGTGACGGCAGAGGTGCAGCCCACTGTGGAGGGGTGTGGGATT 1876
QY 652 SerSerGlyThrGluAsnProHisGlyValLysAlaValThrArgGlyGlnArgCysAla 671
DB 1877 TCTTCTGGCACTGAGAAACCCACATGGAGTGAAGCTGTCCACAGGGGGCAGCGCTGCGCC 1936
QY 672 IleAlaLeuTyrPheThrLeuAspProArgHisSerGluArgAspArgValGlnAlaAsp 691
DB 1937 ATCGCCCTGTGGTTCACGCTGGATCCTCGGCACAGTGCAGAGACAGAGAGAGAGAT 1996
QY 692 AspLeuValLysMetLeuPheSerProGluGlu 702
DB 1997 GACCTGTGTGAAGATGCTGTTTCAGCCCAAGAGAG 2029

RESULT 14
US-10-045-815-5
; Sequence 5, Application US/10045815
; Publication NO. US20020160498A1
; GENERAL INFORMATION:
; APPLICANT: Wadhwah, Renu
; APPLICANT: Sugihara, Takashi
; APPLICANT: Ohide, Akiko
; TITLE OF INVENTION: TUMOR SUPPRESSOR GENE
; FILE REFERENCE: 06501-091001
; CURRENT APPLICATION NUMBER: US/10/045,815
; CURRENT FILING DATE: 2001-10-26
; PRIOR APPLICATION NUMBER: PCT/JP00/02731
; PRIOR FILING DATE: 2000-04-26
; PRIOR APPLICATION NUMBER: JP 11/118806
; PRIOR FILING DATE: 1999-04-26
; NUMBER OF SEQ ID NOS: 15
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 5
; LENGTH: 2416
; TYPE: DNA
; ORGANISM: Mus musculus
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (12)...(2252)
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 2376_
; OTHER INFORMATION: n = A,T,C or G
US-10-045-815-5

Alignment Scores:
Pred. No.: 1.33e-83
Score: 91.00
Percent Similarity: 100.00%
Best Local Similarity: 100.00%
Query Match: 12.36%
DB: 14

; PRIORITY: 1
; NAME/KEY: misc_feature
; LOCATION: 2282
; OTHER INFORMATION: n = A,T,C or G
US-10-045-815-7

Alignment Scores:
Pred. No.: 1.29e-83
Score: 91.00
Percent Similarity: 100.00%
Best Local Similarity: 100.00%
Query Match: 12.36%
DB: 14

US-10-045-815-4 (1-736) x US-10-045-815-7 (1-2322)
QY 612 TyrSerAlaIleLeuTyrLeuAsnGlyAspPheAspClyVlyAsnPhetYrPheThrGlu 631
DB 1757 TACAGCGCATCTTTTACCTCAATGGCGACTTTCATGGAGAACTTTTACTTTCACAGAA 1816
QY 632 LeuAspAlaLysThrValThrAlaGluValGlnProGlnCysGlyArgAlaValGlyPhe 651
DB 1817 CTAGATGCCAAGACTGTGACGGCAGAGGTGCAGCCCACTGTGGAGGGGTGTGGGATT 1876
QY 652 SerSerGlyThrGluAsnProHisGlyValLysAlaValThrArgGlyGlnArgCysAla 671
DB 1877 TCTTCTGGCACTGAGAAACCCACATGGAGTGAAGCTGTCCACAGGGGGCAGCGCTGCGCC 1936
QY 672 IleAlaLeuTyrPheThrLeuAspProArgHisSerGluArgAspArgValGlnAlaAsp 691
DB 1937 ATCGCCCTGTGGTTCACGCTGGATCCTCGGCACAGTGCAGAGACAGAGAGAGAGAT 1996
QY 692 AspLeuValLysMetLeuPheSerProGluGlu 702
DB 1997 GACCTGTGTGAAGATGCTGTTTCAGCCCAAGAGAG 2029

RESULT 14
US-10-045-815-5
; Sequence 5, Application US/10045815
; Publication NO. US20020160498A1
; GENERAL INFORMATION:
; APPLICANT: Wadhwah, Renu
; APPLICANT: Sugihara, Takashi
; APPLICANT: Ohide, Akiko
; TITLE OF INVENTION: TUMOR SUPPRESSOR GENE
; FILE REFERENCE: 06501-091001
; CURRENT APPLICATION NUMBER: US/10/045,815
; CURRENT FILING DATE: 2001-10-26
; PRIOR APPLICATION NUMBER: PCT/JP00/02731
; PRIOR FILING DATE: 2000-04-26
; PRIOR APPLICATION NUMBER: JP 11/118806
; PRIOR FILING DATE: 1999-04-26
; NUMBER OF SEQ ID NOS: 15
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 5
; LENGTH: 2416
; TYPE: DNA
; ORGANISM: Mus musculus
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (12)...(2252)
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 2376_
; OTHER INFORMATION: n = A,T,C or G
US-10-045-815-5

Alignment Scores:
Pred. No.: 1.33e-83
Score: 91.00
Percent Similarity: 100.00%
Best Local Similarity: 100.00%
Query Match: 12.36%
DB: 14

; PRIORITY: 1
; NAME/KEY: misc_feature
; LOCATION: 2282
; OTHER INFORMATION: n = A,T,C or G
US-10-045-815-7

Alignment Scores:
Pred. No.: 1.29e-83
Score: 91.00
Percent Similarity: 100.00%
Best Local Similarity: 100.00%
Query Match: 12.36%
DB: 14

US-10-045-815-4 (1-736) x US-10-045-815-7 (1-2322)
QY 612 TyrSerAlaIleLeuTyrLeuAsnGlyAspPheAspClyVlyAsnPhetYrPheThrGlu 631
DB 1757 TACAGCGCATCTTTTACCTCAATGGCGACTTTCATGGAGAACTTTTACTTTCACAGAA 1816
QY 632 LeuAspAlaLysThrValThrAlaGluValGlnProGlnCysGlyArgAlaValGlyPhe 651
DB 1817 CTAGATGCCAAGACTGTGACGGCAGAGGTGCAGCCCACTGTGGAGGGGTGTGGGATT 1876
QY 652 SerSerGlyThrGluAsnProHisGlyValLysAlaValThrArgGlyGlnArgCysAla 671
DB 1877 TCTTCTGGCACTGAGAAACCCACATGGAGTGAAGCTGTCCACAGGGGGCAGCGCTGCGCC 1936
QY 672 IleAlaLeuTyrPheThrLeuAspProArgHisSerGluArgAspArgValGlnAlaAsp 691
DB 1937 ATCGCCCTGTGGTTCACGCTGGATCCTCGGCACAGTGCAGAGACAGAGAGAGAGAT 1996
QY 692 AspLeuValLysMetLeuPheSerProGluGlu 702
DB 1997 GACCTGTGTGAAGATGCTGTTTCAGCCCAAGAGAG 2029

RESULT 14
US-10-045-815-5
; Sequence 5, Application US/10045815
; Publication NO. US20020160498A1
; GENERAL INFORMATION:
; APPLICANT: Wadhwah, Renu
; APPLICANT: Sugihara, Takashi
; APPLICANT: Ohide, Akiko
; TITLE OF INVENTION: TUMOR SUPPRESSOR GENE
; FILE REFERENCE: 06501-091001
; CURRENT APPLICATION NUMBER: US/10/045,815
; CURRENT FILING DATE: 2001-10-26
; PRIOR APPLICATION NUMBER: PCT/JP00/02731
; PRIOR FILING DATE: 2000-04-26
; PRIOR APPLICATION NUMBER: JP 11/118806
; PRIOR FILING DATE: 1999-04-26
; NUMBER OF SEQ ID NOS: 15
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 5
; LENGTH: 2416
; TYPE: DNA
; ORGANISM: Mus musculus
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (12)...(2252)
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 2376_
; OTHER INFORMATION: n = A,T,C or G
US-10-045-815-5

Alignment Scores:
Pred. No.: 1.33e-83
Score: 91.00
Percent Similarity: 100.00%
Best Local Similarity: 100.00%
Query Match: 12.36%
DB: 14

; PRIORITY: 1
; NAME/KEY: misc_feature
; LOCATION: 2282
; OTHER INFORMATION: n = A,T,C or G
US-10-
```

DB: 14 Gaps: 0  
US-10-045-815-4 (1-736) x US-10-045-815-5 (1-2416)  
QY 612 TyrSerAlaIleLeuTyrLeuAsnGlyAspPheAspGlyGlyAsnPheTyrPheThrGlu 631  
Db 1851 TACAGGGCATCCCTTACTCAATGGCAGCTTCGATGGAGAACTTTTACTTCACAGAA 1910  
QY 632 LeuAspAlaLysThrValThrAlaGluValGlnProGlnCysGlyArgAlaValGlyPhe 651  
Db 1911 CTAGATGCCAAGACATGTGAGCGACAGGTGCAGCCCCAGTGTGGAAGGCTGTGGGATTC 1970  
QY 652 SerSerGlyThrGluAsnProHisGlyValLysAlaValThrArgGlyGlnArgCysAla 671  
Db 1971 TCTTCTGCACCTGAGAACCCACATGAGTGAAGGCTGTCCAGGGGGCAGCGTGGGCC 2030  
QY 672 IleAlaLeuTppPheThrLeuAspProArgHisSerGluArgAspArgValGlnAlaAsp 691  
Db 2031 ATCGCCCTGTGTTTACGCTGGATCTCGGCACAGTGAAGAGACAGGGTGCAGGCAGAT 2090  
QY 692 AspLeuValLysMetLeuPheSerProGluGlu 702  
Db 2091 GACCTGGTGAAGATGCTTTCAGCCAGAGAG 2123

## RESULT 15

US-09-918-995-29327  
; Sequence 29327, Application US/09918995  
; Publication No. US20030073623A1  
; GENERAL INFORMATION:  
; APPLICANT: Hyseq, Inc.  
; TITLE OF INVENTION: NOVEL NUCLEIC ACID SEQUENCES OBTAINED  
; FILE REFERENCE: 20411-756  
; CURRENT APPLICATION NUMBER: US/09/918,995  
; CURRENT FILING DATE: 2001-07-30  
; PRIOR APPLICATION NUMBER: US/09/235,076  
; PRIOR FILING DATE: 1999-01-20  
; NUMBER OF SEQ ID NOS: 38054  
; SOFTWARE: FastSeq for Windows Version 3.0  
; SEQ ID NO 29327  
; LENGTH: 465  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
; FEATURE:  
; NAME/KEY: misc feature  
; LOCATION: (1)..(465)  
; OTHER INFORMATION: n = A,T,C or G  
US-09-918-995-29327

## Alignment Scores:

Pred. No.:	4.16e-80	Length:	465
Score:	87.00	Matches:	87
Percent Similarity:	100.00%	Conservative:	0
Best Local Similarity:	100.00%	Mismatches:	0
Query Match:	11.82%	Indels:	0
DB:	10	Gaps:	0

US-10-045-815-4 (1-736) x US-09-918-995-29327 (1-465)

QY 650 GlyPheSerSerGlyThrGluAsnProHisGlyValLysAlaValThrArgGlyGlnArg 669  
Db 113 GGATTCCTTCAGGCATCGAAACCCACATGGAGTGAAGGCTGTACACGGGGGCGCGC 172  
QY 670 CysAlaIleAlaLeuTppPheThrLeuAspProArgHisSerGluArgAspArgValGln 689  
Db 173 TGTGCCATCGCCCTGTGGTTACCCCTGCACCCCTCGACACAGCGAGCGGAGCGGTGCAG 232  
QY 690 AlaAspAspLeuValLysMetLeuPheSerProGluGluMetAspLeuSerGlnGluGln 709  
Db 233 GCAGATGACCTGGTGAAGATGCTCTTCAGCCCGAGAGAGATGACCTCTCCAGGAGCAG 292  
QY 710 ProLeuAspAlaGlnGlnGlyProProGluProAlaGlnGluSerLeuSerGlySerGlu 729

Search completed: July 18, 2004, 23:47:46  
Job time : 1127 secs

Db 293 CCCTCGATGCCAGCAGGGCCCCCGAACCTGCACAAGAGTCTCTCTCAGGCAGTGAA 352  
QY 730 SerLysProLysAspGluLeu 736  
Db 353 TCGAAGCCCCAAGGATGAGCTA 373

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - nucleic search, using frame\_plus\_p2n model

Run on: July 18, 2004, 10:00:30 ; Search time 6548 Seconds  
(without alignments)  
3356.531 Million cell updates/sec

Title: US-10-045-815-4  
Perfect score: 3870  
Sequence: 1 MAVRALKLTLLAVVAARS.....PPPEAQSLSGSESKPKDEL 736

Scoring table: BLOSUM62

Xgapop 10.0 , Xgapext 0.5  
Ygapop 10.0 , Ygapext 0.5  
Fgapop 6.0 , Fgapext 7.0  
Delop 6.0 , Delext 7.0

Searched: 27513289 seqs, 14931090276 residues

Total number of hits satisfying chosen parameters: 55026578

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:

-MODEL=frame+ p2n model -DRV=slp  
-Q/cgn2\_1/USPTO.spool\_p/US10045815/runat\_14072004\_123008\_16825/app\_query.fasta\_1.903  
-DB=EST -QPM=fastap -SUFFIX=std.rst -MINMATCH=0.1 -LOOPEXT=0 -LOOPEXT=0  
-UNITS=bits -START=1 -END=1 -MATRIX=blosum62 -TRANS=human40.cdi -LIST=45  
-DOCAIGN=200 -THR SCORE=pct -THR MAX=100 -THR MIN=0 -ALIGN=15 -MODE=LOCAL  
-OUTFMT=ptc -NORM=ext -HEADSIZE=500 -MINLEN=0 -MAXLEN=2000000000  
-USER=US10045815 @CGN 1.1 6425 @runat\_14072004\_123008\_16825 -NCPU=6 -ICPU=3  
-NO MMAP -LARGEOUT -NEG SCORE=0 -WAIT -DSPLOCK=100 -LONGLOG  
-DEV TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6  
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :

EST:  
1: em\_estba.\*  
2: em\_esthum.\*  
3: em\_estin.\*  
4: em\_estnu.\*  
5: em\_estov.\*  
6: em\_estpl.\*  
7: em\_estro.\*  
8: em\_hc.\*  
9: gb\_est1.\*  
10: gb\_est2.\*  
11: gb\_hc.\*  
12: gb\_est3.\*  
13: gb\_est4.\*  
14: gb\_est5.\*  
15: em\_estfun.\*  
16: em\_estom.\*  
17: em\_gss\_hum.\*  
18: em\_gss\_inv.\*  
19: em\_gss\_pln.\*  
20: em\_gss\_vrt.\*  
21: em\_gss\_fun.\*  
22: em\_gss\_mam.\*  
23: em\_gss\_mus.\*  
24: em\_gss\_pro.\*  
25: em\_gss\_rod.\*  
26: em\_gss\_pbg.\*  
27: em\_gss\_vrl.\*  
28: gb\_gss1.\*

29: gb\_gss2.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	3819	98.7	2583	11	BC004160	Homo sapi
2	3379	87.3	3186	11	AK030436	Mus muscu
3	3377.5	87.3	2587	11	AK010578	Mus muscu
4	2574	66.5	2211	29	AY411090	Homo sapi
5	2370	61.2	1836	29	AY411092	Mus muscu
6	1685	43.5	1746	29	AY411091	Pan trogl
7	1623	41.9	999	14	CA488534	AGENCOURT
8	1532	39.6	899	13	BU148566	AGENCOURT
9	1510.5	39.0	1101	13	EX364036	EX364036
10	1505	38.9	881	13	EU543499	AGENCOURT
11	1497	38.7	1108	12	EM473346	AGENCOURT
12	1473.5	38.1	1201	13	EX334235	EX334235
13	1459.5	37.7	1201	13	EX442345	EX442345
14	1455	37.6	926	9	AL521774	AL521774
15	1412	36.5	913	12	BG679334	602628511
16	1392	36.0	905	13	BU147415	AGENCOURT
17	1387	35.8	848	13	BU177586	AGENCOURT
18	1360.5	35.2	902	13	EX392014	EX392014
19	1331	34.4	899	13	BU179488	AGENCOURT
20	1321.5	34.1	955	13	BQ678346	AGENCOURT
21	1319	34.1	870	13	BQ676738	AGENCOURT
22	1319	34.1	889	13	BQ678884	AGENCOURT
23	1317.5	34.0	867	14	CF265106	AGENCOURT
24	1298	33.5	939	13	EX353083	EX353083
25	1293	33.4	1201	13	EX445087	EX445087
26	1281	33.1	766	12	BI856252	603382985
27	1276	33.0	929	13	BQ677232	AGENCOURT
28	1264.5	32.7	991	13	BX380701	EX380701
29	1253	32.4	884	13	BU190877	AGENCOURT
30	1253	32.4	915	14	CB196719	AGENCOURT
31	1246.5	32.2	929	13	BQ722593	AGENCOURT
32	1245.5	32.2	941	13	BUB38872	AGENCOURT
33	1244.5	32.2	859	13	BQ213919	AGENCOURT
34	1234	31.6	767	12	BM041968	BM041968
35	1224.5	31.6	1192	12	BX398629	EX398629
36	1210	31.3	959	12	BM802225	AGENCOURT
37	1179.5	30.5	1656	29	AY412908	Homo sapi
38	1173	30.3	758	13	BQ178768	UI-M-EV0-
39	1169	30.2	981	13	BUB39422	AGENCOURT
40	1166	30.1	911	13	BQ720742	AGENCOURT
41	1160.5	30.0	705	10	BE294207	601172737
42	1158.5	29.9	885	13	BQ680853	AGENCOURT
43	1158	29.9	991	13	BQ917813	AGENCOURT
44	1150	29.7	858	14	CD556679	AGENCOURT
45	1148	29.7	655	14	CB552308	MMSF0001_

# ALIGNMENTS

RESULT 1  
BC004160  
LOCUS BC004160  
DEFINITION Homo sapiens leucine proline-enriched proteoglycan (leprecan) 1,  
mRNA (CDNA clone IMAGE:2824480), containing frame-shift errors.  
ACCESSION BC004160  
VERSION BC004160.2 GI:37588922  
KEYWORDS HTC.  
SOURCE Homo sapiens (human)  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
REFERENCE 1 (bases 1 to 2583)





Qy 381 AspValPheGlyIleProPheValAspProAspSerThrProGluGluValIlePro 400  
|  
|  
|  
Db 1171 GATGTTTTGGAAATCCCTTTGGGATCCGGATTCATGGACTCCAGAGGATGATTCC 1230  
|  
|  
|  
Qy 401 LysArgLeuGlnGluLysGlnLysSerGluArgGluThrAlaValArgIleSerGlnGlu 420  
|  
|  
|  
Db 1231 AAGAGATTGCAGAGAAACAGAGAGTCAGACGGGACCCGATCGCATCTCCAGGAG 1290  
|  
|  
|  
Qy 421 IleGlyAsnLeuMetLysGluIleGluThrLeuValGluGluLysThrLysGluSerLeu 440  
|  
|  
|  
Db 1291 ATTGGGAACCTTATGAGGAAATCGAGACCTTTGTGGAAGAGAACCAAGGAGTCACTG 1350  
|  
|  
|  
Qy 441 AspValSerArgIleThrArgGluGlyGlyProLeuLeuTyrGluGlyIleSerLeuThr 460  
|  
|  
|  
Db 1351 GATGTGAGCAGACTGACCCGGGAAGGTGGCCCTGCTGTATGAAGCATCATCTCCACC 1410  
|  
|  
|  
Qy 461 MetAsnSerLysLeuLeuAsnGlyTyrGlnArgValValMetAspGlyValIleSerAsp 480  
|  
|  
|  
Db 1411 ATGAACCTCCAACTCTCTGATGTTCCAGCGGTGTGATGACGCGTAACTCTGAC 1470  
|  
|  
|  
Qy 481 HisGluCysGlnGluLeuGlnArgLeuThrAsnValAlaAlaThrSerGlyAspGlyTyr 500  
|  
|  
|  
Db 1471 CACGAGTGTGAGAGCTGCAGAGACTGACCAATGTGCAGCAACCTCAGGAGATGCTAC 1530  
|  
|  
|  
Qy 501 ArgGlyGlnThrSerProHisThrProAsnGluLysPheTyrGlyValThrValPheLys 520  
|  
|  
|  
Db 1531 CGGGGTGAGACCTCCCACTATCTCCAGCGGTGTGATGACGCGTAACTCTCTCAA 1590  
|  
|  
|  
Qy 521 AlaLeuLysLeuGlnGlnGluGlyLysValProLeuGlnSerAlaHisLeuTyrTyrAsn 540  
|  
|  
|  
Db 1591 GCCTCAAGCTGGGGCAAGGCAAGGTAAGTTCCTCTGCAGAGTCCCACTGTACTACTAAC 1650  
|  
|  
|  
Qy 541 ValThrGluLysValArgAlaGlyMetGluSerTyrPheArgLeuAspThrProLeuTyr 560  
|  
|  
|  
Db 1651 GTACGAGAGAGGTGGCGGCATCATGGATCTCTTCGCGCTGATAGCCCTCTAC 1710  
|  
|  
|  
Qy 561 PheSerTyrSerHisLeuValCysArgThrAlaIleGluGluValGlnAlaGluArgLys 580  
|  
|  
|  
Db 1711 TTTTCTTACTCTCATCTGTGTGTGCGCACTGCCATCGAAGGTCCAGGCAGAGAGGAG 1770  
|  
|  
|  
Qy 581 AspAspSerHisProValHisValAspAsnCysIleLeuAsnAlaGluThrLeuValCys 600  
|  
|  
|  
Db 1771 GATGATAGTATCATCGACGTCGACGCACTGATCTGATCGGAGACCCCTCGTGT 1830  
|  
|  
|  
Qy 601 ValLysGluProProAlaTyrThrPheArgAspTyrSerAlaIleLeuTyrLeuAsnGly 620  
|  
|  
|  
Db 1831 GTCAAGAGCCCCAGCTTACCTTCGCGACTACAGCCCATCTTACCTAAATGG 1890  
|  
|  
|  
Qy 621 AspPheAspGlyLysAsnPheTyrPheThrGluLeuAspAlaLysThrValThrAlaGlu 640  
|  
|  
|  
Db 1891 GACTTCGATGCGGAAACTTTATTTTCACTGAATGATGCCAAGACCCGTGACGCGAGAG 1950  
|  
|  
|  
Qy 641 ValGlnProGlnCysGlyArgAlaValGlyPheSerSerGlyThrGluAsnProHisGly 660  
|  
|  
|  
Db 1951 GTGAGGCTCAGTGTGAAGAGCGGTGGATTTCTTCAGGACTGAAACCCACATGGA 2010  
|  
|  
|  
Qy 661 ValLysAlaValThrArgGlyGlnArgCysAlaIleAlaLeuTrpPheThrLeuAspPro 680  
|  
|  
|  
Db 2011 GTGAAGGCTGTCCACAGGGGCGAGCGTGTGCCATCGCTGTGTGTTACCTGGACCT 2070  
|  
|  
|  
Qy 681 ArgHisSerGluArgAspArgValGlnAlaAspLeuValLysMetLeuPheSerPro 700  
|  
|  
|  
Db 2071 CGACACAGCGAGGGGACAGGGGTGACGAGATGACCTGTGAGATGCTCTTCAGCCCA 2130  
|  
|  
|  
Qy 701 GluGluMetAspLeuSerGlnGluProLeuAspAlaGlnGlnGlyProProGluPro 720  
|  
|  
|  
Db 2131 GAAGAGATGACCTCTCCAGGAGCAGCCCTGGATGCCAGAGGGGCCCCCGAACCT 2190  
|  
|  
|  
Qy 721 AlaGlnGluSerLeuSerGlySerGluSerLysProLysAspGluLeu 736  
|  
|  
|  
Db 2191 GCACAGAGTCTCTCTCAGGAGTGAATCGAAGCCCAAGGATGAGCTA 2238

RESULT 2  
AK030436  
LOCUS  
DEFINITION  
MUS MUSCULUS ADULT MALE PITUITARY gland cDNA, RIKEN full-length  
enriched library, clone:5330411L17 product:leprecan 1, full insert  
sequence.  
AK030436  
VERSION  
AK030436.1 GI:26326436  
KEYWORDS  
HTC; CAP trapper.  
SOURCE  
Mus musculus (house mouse)  
ORGANISM  
Mus musculus  
REFERENCE  
1 Carninci, P. and Hayashizaki, Y.  
High-efficiency full-length cDNA cloning  
Meth. Enzymol. 303, 19-44 (1999)  
99279253  
PUBMED  
10349636  
2  
Carninci, P., Shibata, Y., Hayatsu, N., Sugahara, Y., Shibata, K.,  
Itoh, M., Konno, H., Okazaki, Y., Muramatsu, M. and Hayashizaki, Y.  
Normalization and subtraction of cap-trapper-selected cDNAs to  
prepare full-length cDNA libraries for rapid discovery of new genes  
Genome Res. 10 (10), 1617-1630 (2000)  
20493374  
MEDLINE  
11042159  
3  
Shibata, K., Itoh, M., Aizawa, K., Nagaoka, S., Sasaki, N., Carninci, P.,  
Konno, H., Akiyama, J., Nishi, K., Kitsuura, T., Tashiro, H., Itoh, M.,  
Sumi, N., Ishii, Y., Nakamura, S., Hazama, M., Nishino, T., Harada, A.,  
Yamamoto, R., Matsumoto, H., Sakaguchi, S., Ikegami, T., Kashiwagi, K.,  
Fujiwara, S., Inoue, K., Togawa, Y., Izawa, M., Ohara, E., Watabiki, M.,  
Yoneda, Y., Ishikawa, T., Ozawa, K., Tanaka, T., Matsura, S., Kawai, J.,  
Okazaki, Y., Muramatsu, M., Inoue, Y., Kira, A. and Hayashizaki, Y.  
RIKEN integrated sequence analysis (RISA) system--384-format  
sequencing pipeline with 384 multicapillary sequencer  
Genome Res. 10 (11), 1757-1771 (2000)  
20530913  
PUBMED  
11076861  
4  
The RIKEN Genome Exploration Research Group Phase II Team and the  
FANTOM Consortium.  
Functional annotation of a full-length mouse cDNA collection  
Nature 409, 685-690 (2001)  
5  
The FANTOM Consortium and the RIKEN Genome Exploration Research  
Group Phase I & II Team.  
Analysis of the mouse transcriptome based on functional annotation  
of 60,770 full-length cDNAs  
Nature 420, 563-573 (2002)  
6 (bases 1 to 3186)  
Adachi, J., Aizawa, K., Akimura, T., Arakawa, T., Bono, H., Carninci, P.,  
Fukuda, S., Furuno, M., Hanagaki, T., Hara, A., Hashizume, W.,  
Hayashida, K., Hayatsu, N., Hiramoto, K., Hiraoka, T., Hirozane, T.,  
Hori, F., Imotani, K., Ishii, Y., Itoh, M., Nagawa, I., Kasukawa, T.,  
Kato, H., Kawai, J., Kojima, Y., Kondo, S., Konno, H., Kouda, M.,  
Koya, S., Kurihara, C., Matsuyama, T., Miyazaki, A., Murata, M.,  
Nakamura, M., Nishi, K., Nomura, K., Numazaki, R., Ohno, M., Ohsato, N.,  
Okazaki, Y., Saito, R., Saitoh, H., Sakai, C., Sakai, K., Sakazume, N.,  
Sano, H., Sasaki, D., Shibata, K., Shinagawa, A., Shiraki, T.,  
Sogabe, Y., Tagami, M., Tagawa, A., Takahashi, F., Takaku-Akahira, S.,  
Takeda, Y., Tanaka, T., Tomaru, A., Toya, T., Yasunishi, A.,  
Muramatsu, M. and Hayashizaki, Y.  
Direct Submission  
Submitted (16-JUL-2001) Yoshihide Hayashizaki, The Institute of  
Physical and Chemical Research (RIKEN), Laboratory for Genome  
Exploration Research Group, RIKEN Genomic Sciences Center (GSC),  
RIKEN Yokohama Institute, 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama,  
Kanagawa 230-0045, Japan (E-mail:genome-res@gsc.riken.go.jp,  
URL:http://genome.gsc.riken.go.jp/, Tel:81-45-503-9222,  
Fax:81-45-503-9216)  
CDNA library was prepared and sequenced in Mouse Genome

COMMENT

Encyclopedia Project of Genome Exploration Research Group in Riken  
Genomic Sciences Center and Genome Science Laboratory in RIKEN.  
Division of Experimental Animal Research in Riken contributed to  
prepare mouse tissues.  
Please visit our web site for further details.

## FEATURES

source

```

1. 3186
/organism="Mus musculus"
/mol_type="mRNA"
/strain="C57BL/6J"
/db_xref="FANTOM_DB:5330411L17"
/db_xref="MGI:2393259"
/db_xref="taxon:10090"
/clone="5330411L17"
/sex="male"
/tissue_type="pituitary gland"
/clone_lib="RIXEN full-length enriched mouse cDNA library"
/dev_stage="adult"
44. 2260

```

5

```

1:..:2200
/note="unnamed protein product; leprecan 1
(MGD|GI:188921)
putative"
/codon_start=1
/protein_id="BAC26962.1"
/db_xref="GI:26326437"
/translation="NAVSERRLLAAVLAAVAAALAAVAAESEPGWVAAAPDLLVAEGCT
AAVSGDVGPGVVLNMRALRALPRLSCTRCA TELPWAPDLIDGLDPSLSODP
GAALHDAARFTGAVLRARCLRCGLPPGSAHLLISELDLEFFNKSPYNLYOVAFYKLN
KLEKAAHLLRFTVGNPEHMEWRQNDIYQTMSGVKAEQDRLDEAKPHMEHFEVGRKRL
QYSEKPGAVPHLEALQOEYFVADBEACALCGPYDGYNDLYDSADLPQAITDHYV
VLMKQKQCVATELASHGREKPFDFPLSHYNYLQAYNIGNYQAECAKYLILFF
PNDEVMHQLAYITAMLGBEEAS SISPENAESEYRRERLLEKLLFFAYDIFPFVDY
PDSMTPEEVIKRLQEKQKSEETA VRSQEIIGNLKEIETLVEEKKFESLDVRLRTRP
GGPLLVEGSLTVLKALKIQEGKVPQLQASARMYVNTVKYRVWVESYFLDITPLVYSY
HTPKCEYGLTVLKALKIQEGKVPQLQASARMYVNTVKYRVWVESYFLDITPLVYSY
HLVCFRTALIEEQAERKDSHPHVDCINLNAALMCIKPPAVTFREDYSAILYNGDPL
DQGNFYFOADKVTVAEPQCGRAEFGSTENPHGVKAVTRGCRGASALHMTFLDPL
RQSEDRVVOADLVKXMLFSPVEEDVQAEQPLPDQQGSDEEESLDRGSLXKDEL"
3172..3177
/note="putative"
3186
/note="putative"
signal
site

```

polvA signal

100

3186

/note="putative"

## ORIGIN

Alignment Scores:

Argument Scores:			
Pred. No.:	0	Length:	3186
Score:	3379.00	Matches:	644
Percent Similarity:	91.75%	Conservative:	34
Best Local Similarity:	87.14%	Mismatches:	57
Query Match:	87.31%	Indels:	3
DB:	11	Gaps:	3

US-10-045-815-4 (1-736) x AK030436 (1-3186)

1	MetAlaValArgAlaLeuLysLeuLeuThrThrLeuLeuAlaValAlaAlaAlaSer	20
44	ATGGCGGTGAGCGAAAGAGAGGTCTGGCAGCATCTACTACGGTCCGCGCGCGCGCGCC	103
21	---GlnAlaGluValGluSerGluAlaGlyTyrGlyMetValThrProAspLeuLeuPhe	39
104	TTACCGGTGGCGGAGTCTGAGCCGGATGGAGCTGGAGCCCTGACCTCTCTTAC	163
40	AlaGluGlyThrAlaAlaTyrAlaArgGlyAspTyrProGlyValValLeuSerMetGlu	59
164	GCAGAGGGACCGCGGCTACTTCGCGCGGAGCTGCGCGGGTGGTCTGAACATGGAG	223
60	ArgAlaLeuArgSerArgAlaAlaLeuArgAlaLeuArgLeuArgCysArgThrGlnCys	79
224	CGGCGCTTGGCTCGCGGGCGGCCCTCGCGGCCCTCGGCTCGGCTGCGGCACAGCTGT	283
80	AlaAlaAspPheProTyrGluLeuAspProAspTyrSerProSerPro-----AlaGln	97





```
Db 1545 GATGGTACCGAGGTGACACCTCCACACACCCCAATGAAAGTTCATGGTTACT 1604
Qy 518 ValPheLysAlaLeuLysLeuGlyGlnGluGlyLysValProLeuGlnSerAlaHisLeu 537
Db 1605 GTCTCTCAAGCTCTCAAGCTCGGCGAGGAAGAAAGTTCTCTGCGAGAGTCCCGCATG 1664
Qy 538 TyrTyrAsnValThrGluLysValArgArgIleMetGluSerTyrPheArgLeuAspThr 557
Db 1665 TACTACACGTGCACAGAGGTGCGCGCGTTCATGGAGTCTTCTTCCGCTTGACACG 1724
Qy 558 ProLeuTyrPheSerTyrSerHisLeuValCysArgThrAlaIleGluValGlnAla 577
Db 1725 CCCTCTATTCTCTTATTCCTCCACCTCGTGTGCGGCACTGCAATAGAGAGTCACAGGCT 1784
Qy 578 GluArgLysAspSerHisProValHisValAspAsnCysIleLeuAsnAlaGluThr 597
Db 1785 GAGAGGAGACAGTAGTACGACCCCGTCCAGTGGATACTTCTTCCACAGAACTAGATCCCAAGACTGTG 1844
Qy 598 LeuValCysValLysGluProAlaTyrThrPheArgAspTyrSerAlaIleLeuTyr 617
Db 1845 CTCATGTGTATCAAGGAGGCCCGCCAGCTACACGTTCCGGGACTACAGCGCCCATCTTATC 1904
Qy 618 LeuAsnGlyAspPheAspGlyGlyAsnPheTyrPheThrGluLeuAspAlaLysThrVal 637
Db 1905 CTCATGGCGACTTCGATGAGGAAACTTTTACTTCCACAGAACTAGATCCCAAGACTGTG 1964
Qy 638 ThrAlaGluValGlnProGlnCysGlyArgAlaValGlyPheSerSerGlyThrGluAsn 657
Db 1965 ACGGACAGAGTGCAGCCCGCAGTGTGAAGGGCTGTGGGATTCCTCTTCGCACTGAGAAC 2024
Qy 658 ProHisGlyValLysAlaValThrArgGlyGlnArgCysAlaIleAlaLeuTrpPheThr 677
Db 2025 CCACATGGAGTGAAGCTGTTCACAGGGGGGACGCTGCGCCATCCGCTGTGGTTACG 2084
Qy 678 LeuAspProArgHisSerGluArgAspArgValGlnAlaAspLeuValLysMetLeu 697
Db 2085 CTGATCTCTCGCACAGTAGTACAGAGACAGGCTGACGAGAGTACCTGTGTGAGATGCTG 2144
Qy 698 PheSerProGluGluMetAspLeuSerGlnGluGlnProLeuAspAlaGlnGlyPro 717
Db 2145 TTCAGCCCGAGAGAGGT-GACCTCCCGCAGGAACAGCCCTGCTGACAGCAGGCGCTCG 2203
Qy 718 ProGluProAlaGlnGluSerLeuSerGlySerGluSerLysProLysAspGluLeu 736
Db 2204 CCAGAGCCTGAGAGAGTCTCTGTGACAGAGAGGCTGCTGCACAAAGATGAGCTA 2260

RESULT 4
AY411090
LOCUS
DEFINITION
Homo sapiens LEPREL gene, VIRTUAL TRANSCRIPT, partial sequence,
genomic survey sequence.
ACCESSION
VERSION
AY411090.1
KEYWORDS
GSS.
SOURCE
Homo sapiens (human)
ORGANISM
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE
1 (bases 1 to 2211)
Clark,A.G., Glanowski,S., Nielson,R., Thomas,P., Kejariwal,A.,
Todd,M.A., Tanenbaum,D.M., Civello,D.R., Lu,P., Murphy,B.,
Ferreira,S., Wang,G., Zheng,X.H., White,T.J., Sninsky,J.J.,
Adams,M.D. and Cargill,M.
Inferring nonneutral evolution from human-chimp-mouse orthologous
gene trios
Science 302 (5652), 1960-1963 (2003)
14671302
2 (bases 1 to 2211)
Clark,A.G., Glanowski,S., Nielson,R., Thomas,P., Kejariwal,A.,
Todd,M.A., Tanenbaum,D.M., Civello,D.R., Lu,P., Murphy,B.,
Ferreira,S., Wang,G., Zheng,X.H., White,T.J., Sninsky,J.J.,
Adams,M.D. and Cargill,M.
```

```
TITLE Direct Submission
JOURNAL Submitted (16-NOV-2003) Celera Genomics, 45 West Gude Drive,
Rockville, MD 20850, USA
COMMENT This sequence was made by sequencing genomic exons and ordering
them based on alignment.
FEATURES
source
1..2211
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"
<1..>2211
/gene="LEPREL"
/locus_tag="HCM4108"
ORIGIN
Alignment Scores:
Pred. No.: 3,1e-263 Length: 2211
Score: 2574.00 Matches: 528
Percent Similarity: 71.88% Conservative: 1
Best Local Similarity: 71.74% Mismatches: 207
Query Match: 66.51% Indels: 0
DB: 29 Gaps: 0
```

```
US-10-045-815-4 (1-736) x AY411090 (1-2211)
Qy 1 MetAlaValArgAlaLeuLysLeuLeuThrThrLeuLeuAlaValAlaAlaSer 20
Db 1 ATGGCGGTACGCGGTTGAAGCTGTCACCACTGCTGGCTGCTGGCGGCTGCCTCC 60
Qy 21 GlnAlaGluValGluSerGluAlaGlyTrpGlyMetValThrProAspLeuLeuPheAla 40
Db 61 CAAGCCGAGGTGCGAGTCGAGNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 120
Qy 41 GluGlyThrAlaAlaTyrAlaArgGlyAspTrpProGlyValValLeuSerMetGluArg 60
Db 121 KNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 180
Qy 61 AlaLeuArgSerArgAlaAlaLeuArgAlaLeuArgLeuArgCysArgThrGlnCysAla 80
Db 181 NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 240
Qy 81 AlaAspPheProTyrGluLeuAspProAspTrpSerProSerProAlaGlnAlaSerGly 100
Db 241 NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 300
Qy 101 AlaGlyAlaLeuArgAspLeuSerPhePheGlyGlyLeuLeuArgArgAlaAlaCysLeu 120
Db 301 NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 360
Qy 121 ArgArgCysLeuGlyProProAlaAlaHisSerLeuSerGluGluMetGluLeuPhe 140
Db 361 NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 420
Qy 141 ArgLysArgSerProTyrAsnTyrLeuGlnValAlaTyrPheLysIleAsnLysLeuGlu 160
Db 421 NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 480
Qy 161 LysAlaValAlaAlaAlaHisThrPhePheValGlyAsnProGluHisMetGluMetGln 180
Db 481 AAAGCTGTGTGTCAGCACACACCTTCTTCGTGGGCAATCTCTGAGCAGCATGAAATGAG 540
Qy 181 GlnAsnLeuAspTyrTyrGlnThrMetSerGlyValLysGluAlaAspPheLysAspLeu 200
Db 541 CAGAACCTTAGACTATTACCAACCATGCTCGAGTGAAGGAGGCGGACTTCAAGATCTT 600
Qy 201 GluThrGlnProHisMetGlnGluPheArgLeuGlyValArgLeuTyrSerGluGluGln 220
Db 601 GAGACTCAACCCCATATGCAAGAATTTTCAGTGGAGTGGCGACTCTACTCAGAGAACAG 660
Qy 221 ProGlnGluAlaValProHisLeuGluAlaAlaLeuGlnGluTyrPheValAlaTyrGlu 240
Db 661 CCACAGGAAGCTGTGCCCCCCTACAGCGCGGCTGCAAGATACTTGTGCGCTATGAG 720
```



```
QY 241 GluCysArgAlaLeuCysGluGlyProTyrAspTyrAspGlyTyrAsnTyrLeuGluTyr 260
Db 721 GAGTGGCGTGCCTCTCGAAGGGCCCTATGACTACGATGGCTACACTACCTTGAATAC 780
QY 261 AsnAlaAspLeuPheGlnAlaIleThrAspHisTyrIleGlnValLeuAsnCysLysGln 280
Db 781 AACCTGACCTCTCCAGGCCATCAGATCATATCATCCAGGTCTCACTGTAAGCAG 840
QY 281 AsnCysValThrGluLeuAlaSerHisProSerArgGluLysProPheGluAspPheLeu 300
Db 841 AACTGTGTACGGAGCTTGTCTCCACCAAGTCAGAGAAGCCCTTTGAAGACTTCCCTC 900
QY 301 ProSerHisTyrAsnTyrLeuGlnPheAlaTyrTyrAsnIleGlyAsnTyrThrGlnAla 320
Db 901 CCATCGCATATATATCTGCAGTTTGCCTACTATCAACATTGGGAATATATACAGGCT 960
QY 321 GlyGluCysAlaTyrThrTyrLeuLeuPhePheProAsnAspGluValMetAsnGlnAsn 340
Db 961 GTTGAATGTGCCAAGACTATCTTCTTCTCCCAATGACGAGGTGATGAACCAAAAT 1020
QY 341 LeuAlaTyrTyrAlaAlaMetLeuGlyGluGluHisThrArgSerIleGlyProArgGlu 360
Db 1021 TTGGCTTATTATGAGCTATGCTTGGAGAAGAACACACAGATCCATCGGCCCCCGTGAG 1080
QY 361 SerAlaLysGluTyrArgGlnArgSerLeuLeuGluLysGluLeuLeuPhePheAlaTyr 380
Db 1081 NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 1140
QY 381 AspValPheGlyIleProPheValAspProAspSerTrpThrProGluGluValIlePro 400
Db 1141 NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 1200
QY 401 LysArgLeuGlnGluLysGlnLysSerGluArgGluThrAlaValArgIleSerGlnGlu 420
Db 1201 AAGAGATTGCAGAGAAACAGAGACTCAGACGGGAACACCGGTACGCATCTCCAGGAG 1260
QY 421 IleGlyAsnLeuMetLysGluIleGluThrLeuValGluGluLysThrLysGluSerLeu 440
Db 1261 ATTGGGAACCTTATGAAGAAATCGACACCTTGTGGAGAGAAGACCAAGGAGTCACTG 1320
QY 441 AspValSerArgLeuThrArgGluGlyGlyProLeuLeuTyrGluGlyIleSerLeuThr 460
Db 1321 GATGTGACGACACTGACCCGGGAAGTGGCCCTCTGCTGTATGAAGGCATCAGTCTCACC 1380
QY 461 MetAsnSerLysLeuLeuAsnGlyTyrGlnArgValValMetAspGlyValIleSerAsp 480
Db 1381 ATGAACCTCAAACCTCTCTGAATGTTCCACGGGTGTGTGATGACGGCGTAACTCTGTAC 1440
QY 481 HisGluCysGlnGluLeuGlnArgLeuThrAsnValAlaIleThrSerGlyAspGlyTyr 500
Db 1441 CACGAGTGTGAGGAGTGCAGAGACTACCAATGTGGCAGCAACCTCAGGAGATGGCTAC 1500
QY 501 ArgGlyGlnThrSerProHisThrProAsnGluLysPheTyrGlyValThrValPheLys 520
Db 1501 CGGGGTGAGACTCCCCACATCTCCCAATGAAGAAGTCTATGTTGTCTACTGCTTTCAA 1560
QY 521 AlaLeuLysLeuGlyGlnGluLysValProLeuGlnSerAlaHisLeuTyrTyrAsn 540
Db 1561 GCCCTCAAGCTGGGGCAAGAGGCAAAAGTTCCTCTGCAGAGTCCCACTGTACTACAAC 1620
QY 541 ValThrGluLysValArgArgIleMetGluSerTyrPheArgLeuAspThrProLeuTyr 560
Db 1621 GTGACGAGAGAGTGGCGGCATCATAGACTCTACTTCCGCTGTATACGCCCTCTCTAC 1680
QY 561 PheSerTyrSerHisLeuValCysArgThrAlaIleGluValGlnAlaGluArgLys 580
Db 1681 TTTTCTTACTCTCATCTGGTGTGGCCACTGCCATCGCAAGAGGTCCAGGCGAGAGGAAG 1740
QY 581 AspAspSerHisProValHisValAspAsnCysIleLeuAsnAlaGluThrLeuValCys 600
Db 1741 GATGATAGTATCATCGTCCAGTGGACACTGATCTGATGCCAGACCCCTCGTGTGT 1800
QY 601 ValLysGluProAlaTyrThrPheArgAspTyrSerAlaIleLeuTyrLeuAsnGly 620
```

```
Db 1801 GTCAAGAGCCCCAGCTACACCTTCGCGACTACAGCGCATCTTTACTTAATGGG 1860
QY 621 AspPheAspGlyGlyAsnPheTyrPheThrGluLeuAspAlaLysThrValThrAlaGlu 640
Db 1861 GACTTCGATGCGGAAACTTTTATTCTACTGAATGATGCAAGACCGTGACGNNNNN 1920
QY 641 ValGlnProGlnCysGlyArgAlaValGlyPheSerSerGlyThrGluAsnProHisGly 660
Db 1921 NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 1980
QY 661 ValLysAlaValThrArgGlyGlnArgCysAlaIleAlaLeuTrpPheThrLeuAspPro 680
Db 1981 NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 2040
QY 681 ArgHisSerGluArgAspArgValGlnAlaAspAspLeuValLysMetLeuPheSerPro 700
Db 2041 NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 2100
QY 701 GluGluMetAspLeuSerGlnGluGlnProLeuAspAlaGlnGlnGlyProGluPro 720
Db 2101 GAAGAGATGGACCTCTCCAGAGCAGCCCTGATGCCAGCAGGGCCCCCGAACCT 2160
QY 721 AlaGlnGluSerLeuSerGlySerGlySerLysProLysAspGluLeu 736
Db 2161 GCACAAGAGTCTCTCTCAGGCAGTGAATCGAAGCCCAAGGATGAGCTA 2208

RESULT 5
AY411092 1836 bp DNA linear GSS 16-DEC-2003
LOCUS Mus musculus LEPRE1 gene, VIRTUAL TRANSCRIPT, partial sequence,
DEFINITION genomic survey sequence.
ACCESSION AY411092
VERSION AY411092.1 GI:39767060
KEYWORDS GSS.
SOURCE Mus musculus (house mouse)
ORGANISM Mus musculus
REFERENCE 1 (bases 1 to 1836)
AUTHORS Clark,A.G., Glanowski,S., Nielson,R., Thomas,P., Kejarival,A.,
Todd,M.A., Tanenbaum,D.M., Civello,D.R., Lu,F., Murphy,B.,
Ferriera,S., Wang,G., Zheng,X.H., White,T.J., Shinsky,J.J.,
Adams,M.D. and Cargill,M.
TITLE Inferring nonneutral evolution from human-chimp-mouse orthologous
gene trios
JOURNAL Science 302 (5652), 1960-1963 (2003)
PUBMED 14671302
REFERENCE 2 (bases 1 to 1836)
AUTHORS Clark,A.G., Glanowski,S., Nielson,R., Thomas,P., Kejarival,A.,
Todd,M.A., Tanenbaum,D.M., Civello,D.R., Lu,F., Murphy,B.,
Ferriera,S., Wang,G., Zheng,X.H., White,T.J., Shinsky,J.J.,
Adams,M.D. and Cargill,M.
TITLE Direct Submission
JOURNAL Submitted (16-NOV-2003) Celera Genomics, 45 West Gude Drive,
Rockville, MD 20850, USA
COMMENT This sequence was made by sequencing genomic exons and ordering
them based on alignment.
FEATURES
source
1..1836
/organism="Mus musculus"
/db_type="genomic DNA"
/db_xref="taxon:10090"
<1..>1836
/gene="LEPRE1"
/locus_tag="HCM4108"

gene
Alignment Scores:
Pred. No.: 1,46e-241 Length: 1836
Score: 2370.00 Matches: 462
Percent Similarity: 82.10% Conservative: 15
Best Local Similarity: 79.52% Mismatches: 104
```





```

2 (bases 1 to 1746)
REFERENCE
AUTHORS Clark,A.G., Glanowski,S., Nielson,R., Thomas,P., Kejarawal,A.,
Todd,M.A., Tanenbaum,D.M., Civello,D.R., Lu,F., Murphy,B.,
Ferrera,S., Wang,G., Zheng,X.H., White,T.J., Sniinsky,J.J.,
Adams,M.D. and Cargill,M.
TITLE Direct Submission
JOURNAL Submitted (16-NOV-2003) Celera Genomics, 45 West Gude Drive,
Rockville, MD 20850, USA
COMMENT This sequence was made by sequencing genomic exons and ordering
them based on alignment.
FEATURES
source Location/Qualifiers
1..1746
/organism="Pan troglodytes"
/mol_type="genomic DNA"
/db_xref="taxon:9598"
<1..>1746
/gene="LEPRE1"
/locus_tag="HCM4108"
ORIGIN
Alignment Scores:
Pred. No.: 2,29e-168 Length: 1746
Score: 1695.00 Matches: 356
Percent Similarity: 61.45% Conservative: 1
Best Local Similarity: 61.27% Mismatches: 224
Query Match: 43.54% Indels: 0
DB: 29 Gaps: 0
US-10-045-815-4 (1-736) x AY411091 (1-1746)
QY 156 ILeAsnLysLeuGluLysAlaValAlaAlaHisThrPheValGlyAsnProGlu 175
Db 1 ATCAACAAGTTGGAGAAAGCTGTTCGTGCAGCACACACCTTCTTCGTGGCAATCCTGAG 60
QY 176 HisMetGluMetGlnGlnAsnLeuAspTyrTyrGlnThrMetSerGlyValLysGluAla 195
Db 61 CACATGGAATGGCGAGAACCTTAGACTATTACAAACCATGTCTGGAGCGAGGAGGCC 120
QY 196 AspPheLysAspLeuGluThrGlnProHisMetGlnGluPheArgLeuValArgLeu 215
Db 121 GACTTCAAGGATCTTGAGATCAACCCCATATGCAAGAAATTCGACTGGGAGTGGGACTC 180
QY 216 TyrSerGluGlnProGlnGluAlaValProHisLeuGluAlaAlaLeuGlnGluTyr 235
Db 181 TACTCNAGGAACAGCCACAGGAAGCTGTGCCCATCTAGAGCGCGCTGCAAGAATAC 240
QY 236 PheValAlaTyrGluGluCysArgAlaLeuCysGluGlyProTyrAspTyrAspGlyTyr 255
Db 241 TTGTGGCTATGAGGAGTCCCTGCGCTCGAAGGCGCCCTATGACTGANGATGGCTAC 300
QY 256 AsnTyrLeuGluTyrAsnAlaAspLeuPheGlnAlaAlaLeuThrAspHisTyrLeuGlnVal 275
Db 301 AACTACCTTGATGACAAACGCTGACCTCTCCAGGCCATCACAGNNNNNNNNNNNNNN 360
QY 276 LeuAsnCysLysGlnAsnCysValThrGluLeuAlaSerHisProSerArgGluLysPro 295
Db 361 NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 420
QY 296 PheGluAspPheLeuProSerHisTyrAsnTyrLeuGlnPheAlaTyrTyrAsnLeuGly 315
Db 421 NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 480
QY 316 AsnTyrThrGlnAlaGlyGluCysAlaLysThrTyrLeuLeuPhePheProAsnAspGlu 335
Db 481 AATTANNNNNAGGCTGTGAAUGTGCCAANNCTATCTCTTCCTCCCAATGACGAG 540
QY 336 ValMetAsnGlnAsnLeuAlaTyrTyrAlaAlaMetLeuGluGluHisThrArgSer 355
Db 541 GTGATGAACCAAAATTTGNNCTATTATGACGATATGCTTGGAGAGAAACACACCAAGTCC 600
QY 356 IleGlyProArgGluSerAlaLysGluTyrArgGlnArgSerLeuLeuGluLysGluLeu 375
Db 601 ATCGGCCCCCNCNNCNCNCAAGAGTCTCTNNNAGGCGAGTGAANNAGCCCAAGGATGAG 1740

```

```

QY 376 LeuPhePheAlaTyrAspValPheGlyLeuProPheValAspProAspSerTyrThrPro 395
Db 661 NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 720
QY 396 GluGluValIleProLysArgLeuGlnGluLysGlnLysSerGluArgGluThrAlaVal 415
Db 721 GAAGAGTGTATCCACAGAGATTGNAGAGAAAACAGAAANNNNNNNNNNNNNNNNNNNN 780
QY 416 ArgIleSerGlnGluIleGlyAsnLeuMetLysGluLeuGluThrLeuValGluGluLys 435
Db 781 NNNNNNNCCAGGAGATTGGGACCTTATGAAGAAANNNGANNCCCTTGTGGAAGAAAN 840
QY 436 ThrLysGluSerLeuAspValSerArgLeuThrArgGluGlyGlyProLeuLeuTyrGlu 455
Db 841 NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 900
QY 456 GlyIleSerLeuThrMetAsnSerLysLeuLeuAsnGlyTyrGlnArgValValMetAsp 475
Db 901 NNNNNNNNNCTCACCATGAACCTCCAACTCGTGAATGGTCCAGCGGGTGTGATGCAC 960
QY 476 GlyValIleSerAspHisGluCysGlnGluLeuGlnArgLeuThrAsnValAlaAlaThr 495
Db 961 GGCGTAATCTCTGACCANNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 1020
QY 496 SerGlyAspGlyTyrArgGlyGlnThrSerProHisThrProAsnGluLysPheTyrGly 515
Db 1021 NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 1080
QY 516 ValThrValPheLysAlaLeuLysLeuGlyGlnGluLysValProLeuGlnSerAla 535
Db 1081 NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 1140
QY 536 HisLeuTyrTyrAsnValThrGluLysValArgArgIleMetGluSerTyrPheArgLeu 555
Db 1141 CACTGTACTACACGTGACGAGAGAGTGGCGGCATCATGGAGTCTCTACTTCCGCTG 1200
QY 556 AspThrProLeuTyrPheSerTyrSerHisLeuValCysArgThrAlaIleGluGluVal 575
Db 1201 GATACGCCCTCTACTTTCTCTCTCATCTGTGTGCGGCATGCGCATCGAAGAGGTC 1260
QY 576 GlnAlaGluArgLysAspAspSerHisProValHisValAspAsnCysIleLeuAsnAla 595
Db 1261 CAGCAGAGAGAGAGAGATGATAGTCATCCAGTCCACNTGGCAGCAACTGCATCTTGAATGCC 1320
QY 596 GluThrLeuValCysValLysGluProProAlaTyrThrPheArgAspTyrSerAlaIle 615
Db 1321 GAGACCTCGTGTGTGTCAAGAGCCCCAGCCCTACACGTTCCGGGACTACAGCGCCATC 1380
QY 616 LeuTyrLeuAsnGlyAspPheAspGlyGlyAsnPheTyrPheThrGluLeuAspAlaLys 635
Db 1381 CTTNNNCTAAATGGGACTTCGATGGCGAAACCTTTATTTCATCTGAACGGATGCCAAG 1440
QY 636 ThrValThrAlaGluValGlnProGlnCysGlyArgAlaValGlyPheSerSerGlyThr 655
Db 1441 ACCGTGACGNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 1500
QY 656 GluAsnProHisGlyValLysAlaValThrArgGlyGlnArgCysAlaIleAlaLeuTyrP 675
Db 1501 NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 1560
QY 676 PheThrLeuAspProArgHisSerGluArgAspArgValGlnAlaAspAspLeuValLys 695
Db 1561 NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 1620
QY 696 MetLeuPheSerProGluGluMetAspLeuSerGlnGlnProLeuAspAlaGlnGln 715
Db 1621 ATGCTCTTCAGCCAGAGAGATGGACCTCTCCAGGAGCAGCCCTGGATGCCAGCAG 1680
QY 716 GlyProProGluProAlaGlnGluSerLeuSerGlySerGluSerLysProLysAspGlu 735
Db 1681 GGCCCCCCCCNNCNCNCAAGAGTCTCTNNNAGGCGAGTGAANNAGCCCAAGGATGAG 1740

```

```

QY      736 Leu 736
Db      1741 CTA 1743

RESULT 7
CA488534
LOCUS   AGENCOURT_10809126 MAPcL Homo sapiens cDNA clone IMAGE:6720300 5',
DEFINITION mRNA sequence.
ACCESSION CA488534
VERSION   CA488534.1 GI:24950669
KEYWORDS EST.
SOURCE    Homo sapiens (human)
ORGANISM  Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 999)
AUTHORS   NIH-MGC http://mgi.nci.nih.gov/.
TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL   Unpublished (1999)
COMMENT    Contact: Robert Strausberg, Ph.D.
           Email: cgabs@emil.nih.gov
           Tissue Procurement: Kristi A. Eglund, Ira Pastan
           CDNA Library Preparation: Invitrogen Corp
           DNA Sequencing by: The I.M.A.G.E. Consortium (LLNL)
           Clone distribution: Agencourt Bioscience Corporation
           found through the I.M.A.G.E. Consortium/LLNL at:
           http://image.llnl.gov
           Plate: L14M14279 row: j column: 12
           High quality sequence stop: 734.
           Location/Qualifiers
             1..999
               /organism="Homo sapiens"
               /mol_type="mRNA"
               /db_xref="taxon:9606"
               /clone="IMAGE:6720300"
               /cell_line="ZR-75-1, MCF7, SK-BR-3, MDA-MB-231,
               hERT-HMEL, LNCaP"
               /lab_host="EMDH108"
               /clone_lib="MAPcL"
               /notes="Vector: pCMV-SPORT6; Site 1: EcoRV; Site 2: Not I;
               Subtracted with brain, liver, kidney and muscle.
               Directionally cloned. Priming method: oligo-dT. Average
               insert size: 1800 bp. Library amplification: 26,000 fold.
               Kristi A. Eglund, James J. Vincent, Robert Strausberg,
               Bungkok Lee & Ira Pastan: Discovery of new breast
               cancer genes encoding membrane and secreted proteins.
               Manuscript submitted."

FEATURES
source
Alignment Scores:
Pred. No.: 3,77e-162 Length: 999
Score: 1623.00 Matches: 322
Percent Similarity: 97.31% Conservative: 4
Best Local Similarity: 96.12% Mismatches: 6
Query Match: 41.94% Indels: 4
DB: 14 Gaps: 0

US-10-045-815-4 (1-736) x CA488534 (1-999)

QY      277 AsnCysLysGlnAsnCysValThrGluLeuAlaSerHisProSerArgGluLysProPhe 296
Db      2 AACTGTAAGCAGAACTGTGTCCAGGAGTTGCTTCCACCCCAAGTCGAGAGAAGCCCTTT 61
QY      297 GluAspPheLeuProSerHisTyrAsnTyrLeuGlnPheAlaTyrTyrAsnIleGlyAsn 316
Db      62 GAAGACTTCTCCCATCGCAATTATATATCTCAGTTTGCCCTACCTATACATTCGGAAT 121
QY      317 TyrThrGlnAlaGlyGluCysAlaLysThrTyrLeuLeuPhePheProAsnAspGluVal 336
Db      122 TATACACAGGCTGTGAATGTGCCAAGACCTATCTTCTCTTCCCAATGACGAGGTG 181

```

```

QY      337 MetAsnGlnAsnLeuAlaTyrTyrAlaAlaMetLeuGlyGluGluHisThrArgSerIle 356
Db      182 ATGAACCAAAATTTGGCTATTATGACGATATGCTTGAGAGAAGAACACACAGATCCATC 241
QY      357 GlyProArgGluSerAlaLysGluTyrArgGlnArgSerLeuLeuGluLysGluLeuLeu 376
Db      242 GGCCCCCGTGAAGTGGCAAGAGTACGACAGCAAGCCCTACTGGAAAAAGAACTGCTT 301
QY      377 PhePheAlaTyrAspValPheGlyIleProPheValAspProAspSerTyrThrProGlu 396
Db      302 TCTTCGCTTATGATGTTTTTGGAAATCCCTTTGCGATCCGGATTCATGGACTCCAGAA 361
QY      397 GluValIleProLysArgLeuGlnGluLysGlnLysSerGluArgGluThrAlaValArg 416
Db      362 GAAGTGATTCCCAAGAGATTGCAAGAGAAAAGAGAGTGCAGAACGGGAAACAGCCGTACGC 421
QY      417 IleSerGlnGluIleGlyAsnLeuMetLysGluIleGluThrLeuValGluGluLysThr 436
Db      422 ATCTCCAGGAGATTGGAACTTATGAGGAATATCGAGACCTTGTGGAAAGAGAAGACC 481
QY      437 LysGluSerLeuAspValSerArgLeuThrArgGluGlyGlyProLeuLeuTyrGluGly 456
Db      482 AAGGAGTCACTGGATGTGAGCAGACTGACCCCGGAAGGTGGCCCTCTGCTGATGAAGGC 541
QY      457 IleSerLeuThrMetAsnSerLysLeuLeuAsnGlyTyrGlnArgValValMetAspGly 476
Db      542 ATCAGTCTCACCATGAACCTCCAACTCTCTGAATGGTTCCACAGCGGGTGTGATGGACGC 601
QY      477 ValIleSerAspHisGluCysGlnGluLeuGlnArgLeuThrAsnValAlaAlaThrSer 496
Db      602 GTAATCTCTGACCACGAGTGTGAGGAGCTGCAGAGACTGACCAATGTGGCAACCACTCA 661
QY      497 GlyAspGlyTyrArgGlyGlnThrSerProHisThrProAsnGluLysPheTyrGlyVal 516
Db      662 GGAGATGGCTACCGGGGTGACAGCTCCCACTCTCCCAATGAAAGTTCTTATGGTGTCT 721
QY      517 ThrValPheLysAlaLeuLysLeuGlyGlnGluLysValProLeuGlnSerAlaHis 536
Db      722 ACTGTCTTCAAGCC-CtCAAGCTGGGCAAGAGGCAAGTTCTCTCTCAGAGTGGCCAC 780
QY      537 LeuTyrTyrAsnValThrGluLysValArgArgIleMetGluSerTyrPheArgLeuAsp 556
Db      781 CTGTACTACAAAGTGCAGCAGAGAGTGGCGGCGCATCATGGAGTCTCTACTTCGCGCTGGAT 840
QY      557 ThrProLeuTyrPheSerTyrSerHisLeuValCysArgThrAlaIleGluGluVal-Gl 576
Db      841 ACGCCCTCTACTTCTTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 900
QY      576 nAlaGluArgLysAspAspSerHisProValHisValAspAsnCysIleLeuAsn-Alag 596
Db      901 GGCANANAGGAAGGATGATAGTCTTCCAGTCCAGTGGACACTGCATCTCTGAATGGCCA 960
QY      596 luThrLeuValCysVal-LysGluProProAlaTyrThr 608
Db      961 AGACCTCTCTGTGTCAAAAGAGGCGCCAGCTACACN 999

RESULT 8
LOCUS   BUI48566
DEFINITION BUI48566 899 bp mRNA linear EST 03-SEP-2002
5', mRNA sequence.
ACCESSION BUI48566
VERSION   BUI48566.1 GI:22662098
KEYWORDS EST.
SOURCE    Homo sapiens (human)
ORGANISM  Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 899)
AUTHORS   NIH-MGC http://mgi.nci.nih.gov/.
TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL   Unpublished (1999)
COMMENT    Contact: Robert Strausberg, Ph.D.

```

Email: cgapbs-r@mail.nih.gov  
 Tissue Procurement: DCTD/DTP  
 cDNA Library Preparation: Rubin Laboratory  
 cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)  
 DNA Sequencing by: Agencourt Bioscience Corporation  
 Clone distribution: MGC clone distribution information can be  
 found through the I.M.A.G.E. Consortium/LLNL at:  
<http://image.llnl.gov>  
 Plate: LICM2568 row: n column: 15  
 High quality sequence stop: 724.

# FEATURES

Location/Qualifiers  
 1..899  
 /organism="Homo sapiens"  
 /mol\_type="mRNA"  
 /db\_xref="taxon:9606"  
 /clone="IMAGE:6380366"  
 /tissue\_type="carcinoma, cell line"  
 /lab\_host="DH10B (phage-resistant)"  
 /clone\_lib="NIH\_MGC\_40"  
 /note="Organ: prostate; Vector: pOTB7; Site 1: XhoI;  
 Site 2: EcoRI; cDNA made by oligo-dT priming.  
 Directionally cloned into EcoRI/XhoI sites using the  
 following 5' adaptor: GGCACGAG(G). Library constructed by  
 Ling Hong in the laboratory of Gerald M. Rubin (University  
 of California, Berkeley) using ZAP-cDNA synthesis kit  
 (Stratagene) and Superscript II RT (Life Technologies).  
 Note: this is a NIH\_MGC Library."

## ORIGIN

Alignment Scores:  
 Pred. No.: 1.69e-152 Length: 899  
 Score: 1532.00 Matches: 293  
 Percent Similarity: 98.33% Conservative: 1  
 Best Local Similarity: 97.99% Mismatches: 5  
 Query Match: 39.59% Indels: 1  
 DB: 13 Gaps: 0

US-10-045-815-4 (1-736) x BUI48566 (1-899)

QY 143 ArgSerProTyrAsnTyrLeuGlnValAlaTyrPhelysIleAsnLysLeuGluLysAla 162  
 DB 3 CGGAGCCCTTACCACTACCTGCGAGTCCCTTCAAGATCAACAGTTGGAGAAGCT 62  
 QY 163 ValAlaAlaAlaHisThrPhePheValGlyAsnProGluHisMetGluMetGlnGlnAsn 182  
 DB 63 GTTCTGTCGACACACACTTCTTCTGCGGCACTCTGAGCACATGGAATGACGACAGAAC 122  
 QY 183 LeuAspTyrTyrGlnThrMetSerGlyValLysGluAlaAspPheLysAspLeuGluThr 202  
 DB 123 CTAGACTATTACCAACCACTGTTCTGGAGTGAAGAGCCGACATCAAGGATCTTGAGACT 182  
 QY 203 GlnProHisMetGlnGluPheArgLeuGlyValArgLeuTyrSerGluGluGlnProGln 222  
 DB 183 CAACCCCATATGCAGAAATTCGACTGGAGTGGACTCTACTCAGAGGACACGACACAG 242  
 QY 223 GluAlaValProHisLeuGluAlaAlaLeuGlnGluTyrPheValAlaTyrGluGluCys 242  
 DB 243 GAAGCTGTGCCCCACCTAGAGGGCGCGCTGCAAGAATACCTTGTGCGCTATGAGGAGTGC 302  
 QY 243 ArgAlaLeuCysGluGlyProTyrAspTyrAspGlyTyrAsnTyrLeuGluTyrAsnAla 262  
 DB 303 CGTGGCCCTTCCGAGGGCCCTATGACTAGTGGCTAGCACTACTTGAAGTACACAGCT 362  
 QY 263 AspLeuPheGlnAlaIleThrAspHisTyrIleGlnValLeuAsnCysLysGlnAsnCys 282  
 DB 363 GACCTCTCCAGGCGCATCAGATCATATTACATCCAGTCTCTCACTCACTGTAAGCAGAACTGT 422  
 QY 283 ValThrGluLeuAlaSerHisProSerArgGluLysProPheGluAspPheLeuProSer 302  
 DB 423 GTACGGAGGCTTGTCTCCACCCCACTGAGAGAGGCCCTTTGAGAGACTTCTCCCATCG 482  
 QY 303 HisTyrAsnTyrLeuGlnPheAlaTyrTyrAsnIleGlyAsnTyrThrGlnAlaGlyGlu 322

DB 483 CATTATAATTATCTGCAGTTTGCCTACTATTAACATTTGGGAATTATACACAGCGCTGTGAA 542  
 QY 323 CysAlaLysThrTyrLeuLeuPhePheProAsnAspGluValMetAsnGlnAsnLeuAla 342  
 DB 543 TGTGCCAAGACCTATCTTCTTCTTCCCAATGACGAGGTGATGAATCAAAATTTGGCC 602  
 QY 343 TyrTyrAlaAlaMetLeuGlyGluHisThrArgSerIleGlyProArgGluSerAla 362  
 DB 603 TATTATGAGCTATGCTTGGAGAAACACACAGATCCATCGCCGCCCGTGAGAGTGCC 662  
 QY 363 LysGluTyrArgGlnArgSerLeuLeuGluLysGluLeuPhePheAlaTyrAspVal 382  
 DB 663 AAGGAGTACCCAGCAGGAGCTACTGGAAAAGACGCTTTCTTCGTTATGATGTT 722  
 QY 383 PheGlyIleProPheValAspProAspSerTrpThrProGluGluValIleProLysArg 402  
 DB 723 TTTGGAATTCCTTTGTGATCCGATTCATGCACTCCAGAAGAAGTGAATTCCTCAAGAGA 782  
 QY 403 LeuGlnGluLysGlnLysSerGluArgGluThrAlaValArgIleSerGlnGluIleGly 422  
 DB 783 TTGCAAGAGAAACAGAGTCAACCGGAAACAGCCGTACGCATCTCCAGAGAGATGGG 842  
 QY 423 AsnLeuMetLysGluIleGluThrLeuValGluGluLysThrLysGluSerLeuAsp 441  
 DB 843 AACTT-ATGAAGGAATCGAGACCTTGTGGAAGAGAGACAAAGGAGTCACTGGAT 898

RESULT 9  
 BX364036 1101 bp mRNA linear EST 05-MAY-2003  
 BX364036 Homo sapiens B CELLS (RAMOS CELL LINE) COT 25-NORMALIZED  
 Homo sapiens cDNA clone CS0DL008YM13 5-PRIME, mRNA sequence.  
 BX364036  
 ACCESSION BX364036.1 GI:30370766  
 VERSION  
 KEYWORDS EST.  
 SOURCE Homo sapiens (human)  
 ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 1 (bases 1 to 1101)  
 Li, W.B., Gruber, C., Jessee, J., and Polayes, D.  
 Full-length cDNA libraries and normalization  
 Unpublished (2001)  
 COMMENT  
 Contact: Genoscope  
 Genoscope - Centre National de Sequencage  
 BP 191 91006 Evry cedex - France  
 Email: seqref@genoscope.cns.fr, Web : www.genoscope.cns.fr  
 Library was constructed by Life Technologies, a division of  
 Invitrogen. This sequence belongs to sequence cluster 476.r For  
 more information about this cluster, see  
<http://www.genoscope.cns.fr/cgi-bin/cluster.cgi?seq=CS0DL008AG07QPI&cluster=476.r>. Contact :  
 Feng Liang Email : fliang@lifetech.com URL :  
<http://fulllength.invitrogen.com/InvitrogenCorporation1600>  
 Faraday Avenue Genoscope sequence ID : CS0DL008AG07QPI.  
 FEATURES  
 Location/Qualifiers  
 1..1101  
 /organism="Homo sapiens"  
 /mol\_type="mRNA"  
 /db\_xref="taxon:9606"  
 /clone="CS0DL008YM13"  
 /cell\_type="B CELLS (RAMOS CELL LINE) COT 25-NORMALIZED"  
 /cell\_line="RAMOS CELL LINE"  
 /clone\_lib="Homo sapiens B CELLS (RAMOS CELL LINE) COT 25-NORMALIZED"  
 /note="1st strand cDNA was primed with a NotI-oligo (dT) primer. Five prime end enriched, double-strand cDNA was digested with the Not I and EcoR V sites of the pCMVSPORT 6 vector. Library was normalized."

ORIGIN  
 Alignment Scores:  
 Pred. No.: 4.74e-150 Length: 1101  
 Score: 1510.50 Matches: 300

Percent Similarity: 96.73% Conservative: 2  
Best Local Similarity: 96.15% Mismatches: 10  
Query Match: 39.03% Indels: 4  
DB: 13 Gaps: 0

US-10-045-815-4 (1-736) x BX364036 (1-1101)

Qy	5	AlaLeuLysLeuThrThrLeuLeuAlaValVallalaAlaAspGlnAlaGluVal	24
Dd	160	GCTTGAAGCTGTCACCACTGTGCTGGCGCGCTGCCCTCCAAAGCGAGGTC	219
Qy	25	GluSerGluAlaGlyTyrProAspLeuPheAlaGluGlyThrAla	44
Dd	220	GAGTCGAGCAGATGGGCATGTCGCTCTTCGCGAGGGACCGCA	279
Qy	45	AlaTyraArgGlyAspTrpProGlyValValLeuSerMetGluAlaLeuArgSer	64
Dd	280	GCCTACGCGCGGGGACTGGCGCGGCTGCTGAGCATGGAACGGCGCTCGCGTCC	339
Qy	65	ArgAlaAlaLeuArgAlaLeuArgGlyCysArgThrGlnCysAlaAlaAspPhePro	84
Dd	340	CGGGCAGCCCTTCGGCCCTTCGCTGCGCTGCCACCAGTGTCGCGCGACTCCCG	399
Qy	85	TyrGluLeuAspProAspTrpSerProSerProAlaGlnAlaSerGlyAlaGlyAlaLeu	104
Dd	400	TGGAGCTGGACCCGCTGCTCCCGCCAGCGCGCGCTCGCGCGCGCGCGCTG	459
Qy	105	ArgAspLeuSerPhePheGlyLeuLeuArgAlaAlaCysLeuArgArgCysLeu	124
Dd	460	CGGAGCTGARTCTTCGCGCGCGCTTCGCTGCGCTGCTGCTGCGCGCGCTGCTC	519
Qy	125	GlyProProAlaAlaHisSerLeuSerGluGluMetGluLeuGluPheArgLysArgSer	144
Dd	520	GGCGCGCGCGCGCGCTCGCTCAAGAAAATGGAGCTGGAGTTCCGCAACGGAC-	578
Qy	145	ProTyraAsnTyrLeuGlnValAlaTyrPheLysIleAsnLysLeuGluLysAlaValAla	164
Dd	579	CCCTACAACCTACCTGAGCTGCGCTACTTCAAATCAACAAATTGGAGAAAGCTGTGCT	638
Qy	165	AlaAlaHisThrPhePheValGlyAsnProGluHisMetGluMetGlnGlnAsnLeuAsp	184
Dd	639	CGAGCACACACCTTCTTCGTGGCAATCTCGAGCACATGGAAATGCAGAACCTTAGAC	698
Qy	185	TyrTyrGlnThrMetSerGlyValLysGluAlaAspPheLysAspLeuGluThrGlnPro	204
Dd	699	TATTACCAACCATGCTCGAGTGAAGGAGCGCGACTTCAAGGATCTTTGAGACTCAACCC	758
Qy	205	HisMetGlnGluPheArgLeuGlyValArgLeuTyrSerGluGluGlnProGlnGluAla	224
Dd	759	CATATGCAAGAATTCGACTGGAGTGCACCTCTACTCAGAGAACACGCCACAGAGCT	818
Qy	225	ValProHisLeuGluAlaAlaLeuGlnGluTyrPheValAlaTyrGluGluCysArgAla	244
Dd	819	GTGCCCCACCTAGAGCGCGCTGCAAGAATACTTTGTGGCTATGAGGAGTGGCGTGC	878
Qy	245	LeuCysGluCluProTyrAspTyrAspGlyTyrAsnTyrLeuGluTyrAsnAlaAspLeu	264
Dd	879	CTCTGGAAGGGCCCCTATGATACGATGCTACACTTCTGATACACGCTGACCTC	938
Qy	265	PheGlnAlaIleThrAspHisTyrIleGlnValLeuAsnCysLysGlnAsnCysValThr	284
Dd	939	TTCAGGCCATCACAGATCATTAATCCAGGCTCCTCAACTGTAAGCAAGACTGTGTACG	998
Qy	285	GluLeuAlaSerHisProSerArgGluLysProPheGluAspPheLeuProSerHisTyr	304
Dd	999	GAGCTTCTTCCACCCCAAGTCGAGAGACCTTT-GAAGACTTCTC-CCATGCAATTAT	1056
Qy	305	AsnTyrLeuGlnPheAlaTyrTyrAsnIleGlyAsn	316
Dd	1057	AATTATSTGAGTTTGC-TACTATAMATTGGGANT	1091

RESULT 10  
BU543499

QY 243 ArgAlaLeuCysGluGlyProTyrAspTyrAspGlyTyrAsnTyrLeuGluTyrAsnAla 262  
 Db CGTGCCCTCTCGAAGGCGCCATGACTACGATGCTACACTACCTTGAAGTACAGCGT 362  
 QY 263 AspLeuPheGlnAlaIleThrAspHisTyrIleGlnValLeuAsnCysLysGlnAsnCys 282  
 Db GACCTTCTCGAGGCATCAGATCATTTACATCCAGGTCTCACTGTAGCAGAACTGT 422  
 QY 283 ValThrGluLeuAlaSerHisProSerArgGluLysProPheGluAspPheLeuProSer 302  
 Db GTCCAGGAGCTTGCTTCCACCCCAAGTCGAGAGAAGCCCTTTGAAGACTTCTCCCATCG 482  
 QY 303 HisTyrAsnTyrLeuGlnPheAlaTyrTyrAsnIleGlyAsnTyrThrGlnAlaGlyGlu 322  
 Db CATTATAATTATCTGAGTTTGCCTACTATTAACATTGGGAATTATACACAGGCTTTTGA 542  
 QY 323 CysAlaLysThrTyrLeuLeuPhePheProAsnAspGluValMetAsnGlnAsnLeuAla 342  
 Db TGTGCCAAGACCTATCTCTCTTCTCCCAATGACGAGGTGATGATCAATCAAAATTTGGCC 602  
 QY 343 TyrTyrAlaAlaMetLeuGlyGluGluHisThrArgSerIleGlyProArgGluSerAla 362  
 Db TATTATGAGCTATGCTTTGGAGAAGAACACACACAGATCCATCGGCCCGCTGAGAGTGCC 662  
 QY 363 LysGluTyrArgGlnArgSerLeuLeuGluLysGluLeuLeuPhePheAlaTyrAspVal 382  
 Db AAGAGTACCGACAGCAGAGCTACTCTGAAAAGAACTGCTTTCTTCGTTTATGATTT 722  
 QY 383 PheGlyIleProPheValAspProAspSerTyrThrProGluGluValIleProLysArg 402  
 Db TTTTGAATTTCCCTTTGTGGATCCGATTCATGCACTCCAGAGAAGTGATTCCTCAAGAGA 782  
 QY 403 LeuGlnGluLysGlnLysSerGluArgGluThrAlaValArgIleSerGlnGluIle-Gl 422  
 Db TTGCAAGAAACAGAAAGTACAGACGGAAACAGCCGTACGATCTCCAGGAGATTGG 842  
 QY 422 YAsnLeuMet-LysGluIleGlu-ThrLeuValGluGlu 434  
 Db GAACCTTATGAAGAAATCGAAGACCTTTGTGGAAGAC 881

RESULT 11  
 BM473346 1108 bp mRNA linear EST 05-FEB-2002  
 LOCUS AGENCOURT\_6466598 NIH\_MGC\_88 Homo sapiens cDNA clone IMAGE:5561827  
 DEFINITION 5', mRNA sequence.  
 ACCESSION BM473346  
 VERSION BM473346.1 GI:18522388  
 KEYWORDS EST.  
 SOURCE Homo sapiens (human)  
 ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 1 (bases 1 to 1108)  
 NIH-MGC <http://mgi.nci.nih.gov/>  
 National Institutes of Health, Mammalian Gene Collection (MGC)  
 Unpublished (1999)  
 COMMENT  
 Contact: Robert Strausberg, Ph.D.  
 Email: [cgapbs-rc@mail.nih.gov](mailto:cgapbs-rc@mail.nih.gov)  
 Tissue Procurement: ATCC  
 cDNA Library Preparation: Life Technologies, Inc.  
 cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)  
 DNA Sequencing by: Agencourt Bioscience Corporation.  
 Clone distribution: MGC clone distribution information can be  
 found through the I.M.A.G.E. Consortium/LLNL at:  
<http://image.llnl.gov>  
 Plate: L1AM12230 row: d column: 20  
 High quality sequence start: 37  
 High quality sequence stop: 648.  
 High quality sequence stop: 648.  
 FEATURES  
 Location/Qualifiers  
 1..1108  
 /organism="Homo sapiens"  
 /mol\_type="mRNA"

## ORIGIN

## Alignment Scores:

Pred. No.: 1,338-148 Length: 1108  
 Score: 1497.00 Matches: 310  
 Percent Similarity: 91.01% Conservative: 4  
 Best Local Similarity: 89.86% Mismatches: 27  
 Query Match: 38.68% Indels: 6  
 DB: 12 Gaps: 2

US-10-045-815-4 (1-736) x BM473346 (1-1108)

QY 318 ThrGlnAlaGlyGluCysAlaLysThrTyrLeuLeuPhePheProAsnAspGluValMet 337  
 Db 48 ACACAGGCTGTTGAATGTGCCAAGACCTATCTTCTTCTTCCCCAATGACGAGGTGATG 107  
 QY 338 AsnGlnAsnLeuAlaTyrTyrAlaAlaMetLeuGlyGluGluHisThrArgSerIleGly 357  
 Db 108 AACCAAAATTTGGCCCTATTATGCACTATGCTTGAGAAGAACAACACACAGATCCATCGGC 167  
 QY 358 ProArgGluSerAlaLysGluTyrArgGlnArgSerLeuLeuGluLysGluLeuPhe 377  
 Db 168 CCCCCTGAGAGTGCCAAGAGTACCGACAGCGAGCCCTACTGGAAAAGAACTCTTTC 227  
 QY 378 PheAlaTyrAspValPheGlyIleProPheValAspProAspSerTyrThrProGluGlu 397  
 Db 228 TTCGCTTATGATGTTTTTGGAAATTCCTTTGTGATCCGATTCGGAATCCAGAGAA 287  
 QY 398 ValIleProLysArgLeuGlnGluLysGlnLysSerGluArgGluThrAlaValArgIle 417  
 Db 288 GTGATTCACAGAGATTGCCAAGAGACACAGATCAGAACGGGAACACGCGCTACGATC 347  
 QY 418 SerGlnGluIleGlyAsnLeuMetLysGluIleGluThrLeuValGluLysThrLys 437  
 Db 348 TCCCAGGAGATTGGAACTTATGAAGAAATCGAGACCCCTTGTGGAAGAGAACCAAG 407  
 QY 438 GluSerLeuAspValSerArgLeuThrArgGluGlyGlyProLeuLeuTyrGluGlyLe 457  
 Db 408 GATCCTGATGATGACGAGACTGACCCGGAAGTGGCCCCCTCTGTATGAAGGATC 467  
 QY 458 SerLeuThrMetAsnSerLysLeuLeuAsnGlyTyrGlnArgValValMetAspGlyVal 477  
 Db 468 AGTCTCACCATGAATCCAAACTCCTGAATGTTTCCAGCGGGTGGTATGGACGGCGTA 527  
 QY 478 IleSerAspHisGluCysGlnGluLeuGlnArgLeuThrAsnValAlaAlaThrSerGly 497  
 Db 528 ATCTCTGACACGAGTGTGAGGAGCTGACAGAGCTGACCAATGTGGCAGAACCTCAGGA 587  
 QY 498 AspGlyTyrArgGlyGlnThrSerProHisThrProAsnGluLysPheTyrGlyValThr 517  
 Db 588 GATGGTACCGGGGTGAGACCTCCCCACATATCCCCAATGAAAGTTCTATGGTGCTACT 647  
 QY 518 ValPheLysAlaLeuLysLeuGlyGlnGlyLysValProLeuGlnSerAlaHisLeu 537  
 Db 648 GTCTTCAAGCCCTCAAGCTGGGGCAAGAGGCAAAAGTCTCTCTGACAGAGTGGCCACTG 707  
 QY 538 TyrTyrAsnValThrGluLysValArgIleMetGluSerTyrPheArgLeuAspThr 557  
 Db 708 TACTACAAAGTACGAGAGAGGTGCGCGCATCATGAGATCTTCTTCCGCTGGATACG 767  
 QY 558 ProLeuTyrPheSerTyrSerHisLeuValCysArgThrAlaIleGluGluValGlnAla 577  
 Db 768 CCCCTCTACTTCTCTACTCTCTCTCTGTCGCGCACTGCCCATCGAAGAGTCCCGNCAN 827

/db\_xref="taxon:9606"  
 /clone="IMAGE:5561827"  
 /tissue\_type="duodenal adenocarcinoma, cell line"  
 /lab\_host="DH10B (phage-resistant)"  
 /clone\_lib="NIH\_MGC\_88"  
 /note="Organ: small\_intestine; Vector: pCMV-SPORT6;  
 Site 1: Nci; Site 2: Sall; Cloned unidirectionally;  
 oligo-dt primed. Average insert size 1.767 kb. Library  
 enriched for full-length clones and constructed by Life  
 Technologies. Note: this is a NIH\_MGC Library."



```

QY 578 GluA2GlyAspSerHisProValHisValAspAsnCysIleLeu-AsnAlaGluTh 597
Db 828 AGAAGAA-GATCATAGTCATCAGTCCACGTGGACATGCTGAATGCGAAGAC 886
QY 597 rLeuValCysValLysGluProAlaTyrThr-PheArgAspTyrSerAlaIleLeuT 617
Db 887 CTTGCTGTGTCAAAGAGCCCGAGCT---ACACTTCCGCACTACAGGGATCCTTTA 943
QY 617 rLeuAsnGlyAspPheAspGlyGlyAsnPhetYrPhetThrGluLeuAspAlaLysThrv 637
Db 944 CTTAAAGGGGGGACTTCGATGCGGGAACCTTTATTTCCT-GAATCGATGCCAAACCG 1002
QY 637 alThrAlaGluValGlnProGlnCysGlyArgAlaValGlyPheSer---SerGlyThrG 656
Db 1003 TGACGGCAAGGGGCCACCCCGGAGGGGAAAAACCGGGGAACTCTCTTCGGGCTGG 1062
QY 656 luAsnProHis 659
Db 1063 AAACCCCCAC 1073

RESULT 12
BX334235 1201 bp mRNA linear EST 01-MAY-2003
LOCUS BX334235 Homo sapiens PLACENTA COT 25-NORMALIZED Homo sapiens cDNA
DEFINITION clone CSODI001YG09 5-PRIME, mRNA sequence.
ACCESSION BX334235.1 GI:30310318
VERSION BX334235.1
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryote; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 1201)
Li, W.B., Gruber, C., Jessee, J. and Polayes, D.
Full-length cDNA libraries and normalization
Unpublished (2001)
Contact: Genoscope
Genoscope - Centre National de Sequencage
BP 191 91006 EVRY cedex - France
Email: seqref@genoscope.cns.fr, Web : www.genoscope.cns.fr
Library was constructed by life technologies, a division of
Invitrogen. This sequence belongs to sequence cluster 476.r For
more information about this cluster, see
http://www.genoscope.cns.fr/
cgi-bin/cluster.cgi?seq=CSODI001AD05QPI&cluster=476.r. Contact :
Feng Liang Email : fliang@lifetech.com URL :
http://fulllength.invitrogen.com/ invitrogen Corporation 1600
Faraday Avenue Genoscope sequence ID : CSODI001AD05QPI.
FEATURES
Location/Qualifiers
1..1201
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="CSODI001YG09"
/tissue_type="PLACENTA COT 25-NORMALIZED"
/clone_lib="Homo sapiens PLACENTA COT 25-NORMALIZED"
/note="1st strand cDNA was primed with a NotI-oligo(dT)
primer. Five prime end enriched, double-strand cDNA was
digested with Not I and cloned into the Not I and EcoR V
sites of the pCMVSPORT 6 vector. Library was normalized."
ORIGIN
Alignment Scores:
Pred. No.: 4,96e-146 Length: 1201
Score: 1473.50 Matches: 292
Percent Similarity: 96.69% Conservative: 0
Best local Similarity: 96.69% Mismatches: 9
Query Match: 38.07% Indels: 4
DB: 13 Gaps: 0
US-10-045-815-4 (1-736) x BX334235 (1-1201)

```

```

QY 5 AlaLeuLysLeuLeuThrThrLeuLeuAlaValAlaAlaAlaLaserGlnAlaGluVal 24
Db 109 GCGTTGAAGCTGCTGACACACACTGCTGGCTGCTGGCGCGCTCCCAAGCGAGGTC 168
QY 25 GluSerGluAlaGlyTyrGlyMetValThrProaspLeuLeuPheAlaGluGlyThrAla 44
Db 169 GAGTCCGAGGAGGATGGGCGATGCTGACGCTGATCTGCTTTCGCCGAGGGAGACGCA 228
QY 45 AlaTyrAlaArgGlyAspTyrProGlyValValLeuSerMetGluArgAlaLeuArgSer 64
Db 229 GCCTACGCGCGCGGGAGCTGCCCGGGTGTCTGAGCATGGAAACGGGGCGCTCGCTCC 288
QY 65 ArgAlaAlaLeuArgAlaLeuArgLeuArgCysArgThrGlnCysAlaAlaAspPhePro 84
Db 289 CGGGCAGCCCTCCGCGCCCTTCGCTGCTGCGCCACCCAGTGTGCGCGGACTTCCCG 348
QY 85 TrpGluLeuAspProAspTyrPheSerProAlaGlnAlaSerGlyAlaGlyAlaLeu 104
Db 349 TGGGAGCTGGACCCGAGCTGCTCCCGAGCCCGCCAGGCTCGGGCGCGCGCGCTG 408
QY 105 ArgAspLeuSerPhePheGlyGlyLeuLeuArgAlaAlaCysLeuArgArgCysLeu 124
Db 409 CGGACCTGAGCTTCTTCGGGGGCTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 468
QY 125 GlyProAlaAlaHisSerLeuSerGluLeuMetGluLeuGluPheArgLysArgSer 144
Db 469 GGGCGCGCGCGCGCCACTCGCTCAGCGAAGAGATGGAGCTGGAGTTCCGACAGCGGAGC 528
QY 145 ProTyrAsnTyrLeuGlnValAlaTyrPheLysIleAsnLysLeuGluLysAlaValAla 164
Db 529 CCTACAACTACCTGAGCTGCGCTACTTCAAGATCAACAAGTTGGAGAAAGCTGTGCT 588
QY 165 AlaAlaHisThrPhePheValGlyAsnProGluHisMetGluMetGlnGlnAsnLeuAsp 184
Db 589 GCAGCACACACCTTCTTCGTGGCAATCTGAGCACATGGAATGACAGAACCTAGAC 648
QY 185 TyrTyrGlnThrMetSerGlyValLysGluAlaAspPheLysAspLeuGluThrGlnPro 204
Db 649 TATTACCAACCATCTCTGGAGTGAAGAGCGCGACTTCAAGGATCTTGAGACTCAACCC 708
QY 205 HisMetGlnGluPheArgLeuGlyValArgLeuTyrSerGluGluGlnProGlnGluAla 224
Db 709 CATATGCAAGATTTTCGACTGGAGTGGAGCTACTACTCAGAGGAACAGCCAGGAAGCT 768
QY 225 ValProHisLeuGluAlaAlaLeuGlnGluTyrPheValAlaTyrGluGluCysArgAla 244
Db 769 GTGCCCCCACTAGAGCGCGCTGCAAGAATACTTTGTGGCTATGAGAGTGGCGTGC 828
QY 245 LeuCysGluGlyProTyrAspTyrAspGlyTyrAsnTyrLeuGluTyrAsnAlaAspLeu 264
Db 829 CTCCTCGGAAGGGCCCTATGAMTANGATGGCTACAACTACCTTGAGTACAAAGCTGACCTC 888
QY 265 PheGlnAlaIleThrAspHisTyrIleGlnValLeuAsnCysLysGln-AsnCysValTh 284
Db 889 TTCCAGGCCCATCACAGATCAITACATCCAGGTCCTCAACTGTAARCAAGAACTGTGTAC 948
QY 284 rGluLeuAlaSerHisProSerArgGluLysProPheGluAspPheLeuProSerHisTyr 304
Db 949 GGAGCTGTGTTCCAA-CCAAAGTCGAGAGAACCTTTT-GAAGACTTCTT-CCATYGSATTA 1005
QY 304 rAsn 305
Db 1006 TAAT 1009

RESULT 13
BX442345 1201 bp mRNA linear EST 15-MAY-2003
LOCUS BX442345 Homo sapiens FETAL BRAIN Homo sapiens cDNA clone
DEFINITION CSODF029YB14 5-PRIME, mRNA sequence.
ACCESSION BX442345
VERSION BX442345.1 GI:30776018
KEYWORDS EST.
SOURCE Homo sapiens (human)

```



ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE 1 (bases 1 to 1201)  
AUTHORS Li, W.B., Gruber, C., Jessee, J., and Polayes, D.  
TITLE Full-length cDNA libraries and normalization  
JOURNAL Unpublished (2001)  
COMMENT Contact: Genoscope  
Genoscope - Centre National de Sequencage  
BP 191 91006 EVRY cedex - France  
Email: segre@genoscope.cns.fr, Web : www.genoscope.cns.fr  
Library was constructed by Life Technologies, a division of  
Invitrogen. This sequence belongs to sequence cluster 476.r For  
more information about this cluster, see  
http://www.genoscope.cns.fr/  
cgi-bin/cluster.cgi?seq=CS0DF029DA070Pikcluster=476.r. Contact :  
Feng liang Email : fliang@lifetech.com URL : Corporation 1600  
http://fulllength.invitrogen.com/ Invitrogen Corporation 1600  
Faraday Avenue Genoscope sequence ID : CS0DF029DA070P1.  
Location/Qualifiers  
1..1201  
/organism="Homo sapiens"  
/mol\_type="mRNA"  
/db\_xref="taxon:9606"  
/clone="CS0DF029YB14"  
/tissue\_type="FETAL BRAIN"  
/dev\_stage="fetal"  
/clone\_lib="Homo sapiens FETAL BRAIN"  
/note="Organ: brain; Vector: pCMVSPORT 6; 1st strand cDNA  
was primed with a NotI-oligo (AT) primer. Five prime end  
enriched, double-strand cDNA was digested with NotI and  
cloned into the NotI and EcoRV sites of the pCMVSPORT 6  
vector. Library was not normalized."

ORIGIN  
Alignment Scores:  
Pred. No.: 1,568-144 Length: 1201  
Score: 1459.50 Matches: 299  
Percent Similarity: 91.84% Conservative: 5  
Best Local Similarity: 90.33% Mismatches: 22  
Query Match: 37.71% Indels: 8  
DB: 13 Gaps: 2

US-10-045-815-4 (1-736) x BX442345 (1-1201)

QY 1 MetAlaValArgAlaLeuLysLeuLeuThrLeuLeuAlaValAlaAlaSer 20  
Db 89 ATGCGGTACGCGGTTGAAGCTGCTGACACACTGCTGCTGCGTGGCGCTGCTCC 148  
QY 21 GlnAlaGluValGluSerGluAlaGlyTrpGlyMetValThrProAspLeuLeuPheAla 40  
Db 149 CAACCCGAGGTGAGTCCGAGGAGGATGGGATGAGCGCTGATCTGCTTCGCGC 208  
QY 41 GluGlyThrAlaAlaValAlaArgGlyAspTrpProGlyValValLeuSerMetGluArg 60  
Db 209 GAGGGACCGCAGCTACCGCGCGGGGACTGCGCGGGTGGTCTGAGCATGGAACGG 268  
QY 61 AlaLeuArgSerArg-AlaAlaLeuArgAlaLeuArgLeuArgCysArgThrGlnCysAl 80  
Db 269 GCCTGGCTCCCGGTCACCCCTCCCGCCCTTCGCTCGCTGCGGACCCAGGTGTC 328  
QY 80 aAlaAspPheProTrpGluLeuAspProAspTrpSerProSerProAlaGlnAlaSerG1 100  
Db 329 CGCCGACTTCCCGTGGAGTGGACCCGACTGGTCCCCCAGCCGCCGCGCTCGGG 388  
QY 100 ValAlaGlyAlaLeuArgAspLeuSerPhePheGlyGlyLeuLeuArgArgAlaAlaCysLe 120  
Db 389 CGCGCGCGCCCTCGCGACCTGAGCTTCTTCGGGGGGCCCTTCGCGGCGCTGCGCT 448  
QY 120 uArgArgCysLeuGlyProProAlaAlaHisSerLeuSerGluLeuMetGluLeuGluPh 140  
Db 449 GCGCGCGTGCCTCGGCGCGCGCGCCCACTCGCTCAACGAAATAAATFAAGCTGGAGTT 508

QY 140 eArgLysArgSerProTyrAsnTyrLeuGlnValAlaTyrPheLysIleAsnLysLeuG1 160  
Db 509 CCCAAGCGGAC-CCCTACAACCTACMTGCGAGTCGCTACTTCAAGATCAACAAATTGGA 567  
QY 160 uLysAlaValAlaAlaHisThrPhePheValGlyAsnProGluHisMetGluMetG1 180  
Db 568 GAAAGCTGTTGTCGAGCAGACACACCTTCTTCTGGGCATCTCTGAGCATGGAATGCA 627  
QY 180 nGlnAsnLeuAspTyrTyrGlnThrMetSerGlyValLysGluAlaAspPheLysAspLe 200  
Db 628 GCAGAACCTAGACTATTACCAACCATGCTCTGAGTGAAGGAGGCGGACTTCAAGGATCT 687  
QY 200 uGluThrGlnProHisMetGlnGluPheArgLeuGlyValAlaGlyLeuTyrSerGluGluG1 220  
Db 688 TGAGACTCAACCCCATATGCAAGAAATTCGACTGGAGTGGCGACTCTACTCAGAGGAACA 747  
QY 220 nProGlnGluAlaValProHisLeuGluAlaAlaLeuGlnGluTyrPheValAlaTyr-G 240  
Db 748 GCCACAGGAGCTGTGCCACCTAGAGCGCGCTGCAAGAATACTTTTGGCCTATGG 807  
QY 240 LuGluCysArgAlaLeuCysGluGlyProTyrAspTyrAspGlyTyrAsnTyrLeuGluT 260  
Db 808 AGAGTGGCGTCCCTCTCGAAGGCGCTATGACTAGTGGCTACAACTACCTTGGAGT 867  
QY 260 YrAsnAlaAspLeuPheGlnAlaIleThrAspHisTyrIleGlnValLeuAsnCysLysG 280  
Db 868 ACAACGCTGACCTTCCAGGCGCATCACAGATCATTCACGCTCCCACTGTAARC 927  
QY 280 InAsn-CysValThrGluLeuAlaSerHisProSerArgGluLysProPheGluAspPhe 299  
Db 928 AGAACTGTGTCCAGGAGCTGCTKCCACCAAGTCGAGAGAGCCSTTG---ARACTT 984  
QY 300 LeuProSerHisTyrAsnTyrLeuGlnPheAlaTyrTyrAsnIleGlyAsnTyrThrGln 319  
Db 985 CTKCCATCGCATTAATATCTGCGATT-GCYWMTATWATTTGGGATWAAACA--- 1039  
QY 320 AlaGlyGluCysAlaLysThrTyrLeu 328  
Db 1040 GGYGTGAATGKCAARMCTATYTCTY 1066

RESULT 14  
AL521774  
LOCUS  
DEFINITION AL521774 Homo sapiens NEUROBLASTOMA COT 10-NORMALIZED Homo sapiens  
cDNA clone CS0DB003YL11 5-PRIME, mRNA sequence.  
ACCESSION AL521774  
VERSION AL521774.2 GI:31040051  
KEYWORDS EST.  
SOURCE Homo sapiens (human)  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
REFERENCE 1 (bases 1 to 926)  
AUTHORS Li, W.B., Gruber, C., Jessee, J., and Polayes, D.  
TITLE Full-length cDNA libraries and normalization  
JOURNAL Unpublished (2001)  
COMMENT On Feb 13, 2001 this sequence version replaced gi:12785267.  
Contact: Genoscope  
Genoscope - Centre National de Sequencage  
BP 191 91006 EVRY cedex - France  
Email: segre@genoscope.cns.fr, Web : www.genoscope.cns.fr  
Library was constructed by Life Technologies, a division of  
Invitrogen. This sequence belongs to sequence cluster 476.r For  
more information about this cluster, see  
http://www.genoscope.cns.fr/  
cgi-bin/cluster.cgi?seq=CS0DB003CF06QPIcluster=476.r. Contact :  
Feng liang Email : fliang@lifetech.com URL :  
http://fulllength.invitrogen.com/ Invitrogen Corporation 1600  
Faraday Avenue Genoscope sequence ID : CS0DB003CF06QPI.  
Location/Qualifiers  
1..926  
/organism="Homo sapiens"  
/mol\_type="mRNA"

FEATURES  
source



```
Db      180 ACCAGGAGTCACTGGATGTGACGACTGACCCCGGAGGTGGCCCTGCTGTATGAA 239
Qy      456 GlyIleSerLeuThrMetAsnSerLysLeuLeuAsnGlyTyrGlnArgValValMetAsp 475
Db      240 GGCATCAGTCTCACCATACTCCAACTCCCTGAATGGTCCAGCGGGTGGTGTATGGAC 299
Qy      476 GlyValIleSerAspHisGlnCysGlnGluLeuGlnArgLeuThrAsnValAlaAlaThr 495
Db      300 GCGGTAACTCTGACACACAGAGTTCAGGAGCTGCAGAGACTGACCAATGTGGCAGCAACC 359
Qy      496 SerGlyAspGlyTyrArgGlyGlnThrSerProHisThrProAsnGluLysPheTyrGly 515
Db      360 TCAGGAGATGGCTACCGGGGTCAGACCTCCCCACATACTCCCAATGAAAAGTTCTATGGT 419
Qy      516 ValThrValPheLysAlaLeuLysLeuGlyGlnGluGlyLysValProLeuGlnSerAla 535
Db      420 GTCACGTGCTTCAAAGCCCTCAAGCTGGGGCAAGAAGCAAGTTCCCTCTGCAGAGTGCC 479
Qy      536 HisLeuTyrTyrAsnValThrGluLysValArgArgIleMetGluSerTyrPheArgLeu 555
Db      480 CACCTGTACTACACGTGACGAGAGAGGTGGCGGCATCATGGAGTCTTACTTCCGCGCTG 539
Qy      556 AspThrProLeuTyrPheSerTyrSerHisLeuValCysArgThrAlaIleGluGluVal 575
Db      540 GATAGCCCCCTCTACTTTCTCTACTCTCTCATCTGGTGTCCGCACTGCCATCGAAGAGGTC 599
Qy      576 GlnAlaGluArgLysAspSerHisProValHisValAspAsnCysIleLeuAsnAla 595
Db      600 CAGGACAGAGAGAGAGGATGATGATCCAGTCCACGTGGACAACTGCATCCTGAATGCC 659
Qy      596 GluThrLeuValCysValLysGluProProAlaTyrThrPheArgAspTyrSerAlaIle 615
Db      660 GAGACCTCTGCTGTGTCAAAGAGCCGCCAGCCTACACCTTCGCGCACTACAGCGCCATC 719
Qy      616 LeuTyrLeuAsnGlyAspPheAspGlyGlyAsn-PheTyrPheThrGluLeuAspAlaLy 635
Db      720 CTTTACCTAAATGGGGACTTCGATGGCGGAAACTTTTATTTCACCTGAATGGATGCCAA 779
Qy      635 sThrValThrAlaGluValGlnProGlnCysGly-ArgAlaValAlGlyPheSerSer-Gly 654
Db      780 GACCGTGACGCGCAGAGGTGCAGGCTCAGTGTGGAAAGAGCGGTGGGATTTCTTCCAGGC 839
Qy      655 ThrGlu-AsnProHisGlyValLysAlaValThrArgGlyGlnArgCysAlaIleAlaLe 674
Db      840 ACTGAAAGAACCCACATGGAGTGAAGGCTGTACCCAGGGGCGAGGCTGTGC---ATTGCGCT 896
Qy      674 uTrpPheThrLeuAsp 679
Db      897 GTGGTTCCACCTTGAA 912
```

Search completed: July 18, 2004, 17:57:01  
Job time : 6590 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - nucleic search, using frame\_plus\_p2n model

Run on: July 18, 2004, 16:08:41 ; Search time 10733 Seconds  
(without alignments)

2972.186 Million cell updates/sec

US-10-045-815-4

Title: 736

Sequence: 1 MAVRALKLTLLAVVAAS.....PPEPAQESLSGESKPKDEL 736

Scoring table:

OLIGO  
Xgapop 60.0 , Xgapext 60.0  
Ygapop 60.0 , Ygapext 60.0  
Fgapop 6.0 , Fgapext 7.0  
Delop 6.0 , Delext 7.0

Searched: 3470272 seqs, 21671516995 residues

Word size: 1

Total number of hits satisfying chosen parameters: 6934743

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Command line parameters:

-MODEL=frame+ p2n.model -DEV=xlp  
-Q=/cgn2 1/USPTO\_spool\_p/US10045815/runat 14072004 123119 17240/app\_query.fasta\_1.903  
-DB=GenEmbl -OPMT=fastap -USUFFIX=oli.rge -MINMATCH=0.1 -LOOPEXT=0 -LOOPEXT=C  
-UNITS=bits -START=1 -END=-1 -MATRIX=oligo -TRANS=human40.cdi -LIST=45  
-DOCALIGN=200 -THR SCORE=quality -THR MIN=1 -ALIGN=15 -MODE=LOCAL -OUTFMT=ptc  
-NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000  
-USER=US10045815 @CGN 1 1 7406 @runat 14072004 123119 17240 -NCPU=6 -ICPU=3  
-NO MMAP -LARGQUERY -NEG SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG  
-DEV TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=60 -XGAPEXT=60 -FGAPOP=6  
-FGAPEXT=7 -YGAPOP=60 -YGAPEXT=60 -DELOP=6 -DELEXT=7

Database :

GenEmbl.\*

1: gb.ba.\*  
2: gb.htg.\*  
3: gb.in.\*  
4: gb.om.\*  
5: gb.ov.\*  
6: gb.pat.\*  
7: gb.ph.\*  
8: gb.pl.\*  
9: gb.pr.\*  
10: gb.ro.\*  
11: gb.sts.\*  
12: gb.sy.\*  
13: gb.un.\*  
14: gb.vi.\*  
15: em.ba.\*  
16: em.fun.\*  
17: em.hum.\*  
18: em.in.\*  
19: em.mu.\*  
20: em.om.\*  
21: em.or.\*  
22: em.ov.\*  
23: em.pat.\*  
24: em.ph.\*  
25: em.pl.\*  
26: em.ro.\*  
27: em.sts.\*  
28: em.un.\*

29: em.vi.\*  
30: em.htg.hum.\*  
31: em.htg.inv.\*  
32: em.htg.other.\*  
33: em.htg.mus.\*  
34: em.htg.pln.\*  
35: em.htg.rod.\*  
36: em.htg.mam.\*  
37: em.htg.vrt.\*  
38: em.sy.\*  
39: em.htgo.hum.\*  
40: em.htgo.mus.\*  
41: em.htgo.other.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	736	100.0	2600	9 AF097432	AF097432 Homo sapi
2	433	58.8	2563	6 AX136191	AX136191 Sequence
3	433	58.8	2563	6 BD123548	BD123548 Secretary
4	433	58.8	2563	9 AK075418	AK075418 Homo sapi
5	433	58.8	2583	6 AX704765	AX704765 Sequence
6	433	58.8	2585	6 AX877288	AX877288 Sequence
7	433	58.8	2585	6 BD156581	BD156581 Primer fo
8	433	58.8	2585	9 AK027697	AK027697 Homo sapi
9	407	55.3	2524	6 AX877139	AX877139 Sequence
10	407	55.3	2524	6 BD156505	BD156505 Primer fo
11	407	55.3	2524	9 AK027680	AK027680 Homo sapi
12	360	48.9	2829	9 AF097431	AF097431 Homo sapi
13	301	40.9	1562	9 BC015309	BC015309 Homo sapi
14	295	40.1	2993	6 AX880254	AX880254 Sequence
15	295	40.1	2993	6 BD158280	BD158280 Primer fo
16	295	40.1	2993	9 AK027648	AK027648 Homo sapi
17	289	39.3	1173	9 BT007039	BT007039 Homo sapi
18	289	39.3	1173	12 BT007768	BT007768 Synthetic
19	204	27.7	1526	9 AK025841	AK025841 Homo sapi
20	179	24.3	550	6 AX677402	AX677402 Sequence
21	103	14.0	682	6 BD020571	BD020571 Novel gen
22	103	14.0	682	6 BD100509	BD100509 Novel gen
23	101	13.7	652	6 AX136510	AX136510 Sequence
24	101	13.7	652	6 BD123750	BD123750 Secretary
25	101	13.7	738	6 AX868542	AX868542 Sequence
26	101	13.7	738	6 BD148604	BD148604 Primer fo
27	101	13.7	774	6 AX868815	AX868815 Sequence
28	101	13.7	774	6 BD148877	BD148877 Primer fo
29	101	13.7	2524	10 AF087433	AF087433 Rattus no
30	101	13.7	2569	10 BC024047	BC024047 Mus muscu
31	101	13.7	161835	9 AC098484	AC098484 Homo sapi
32	91	12.4	2297	10 AF165164	AF165164 Mus muscu
33	91	12.4	2391	10 AF165163	AF165163 Mus muscu
34	80	10.9	703	6 AX868722	AX868722 Sequence
35	80	10.9	703	6 BD148784	BD148784 Primer fo
36	73	9.9	711	6 BD020572	BD020572 Novel gen
37	73	9.9	711	6 BD100510	BD100510 Novel gen
38	58	7.9	569	6 BD153901	BD153901 Sequence
39	58	7.9	569	6 AX873839	AX873839 Primer fo
40	47	6.4	193219	10 AL606975	AL606975 Mouse DNA
41	47	6.4	194020	2 AC101988	AC101988 Mus muscu
42	47	6.4	225308	2 AC098918	AC098918 Rattus no
43	42	5.7	554	6 AX873755	AX873755 Sequence
44	42	5.7	554	6 BD153817	BD153817 Primer fo
45	37	5.0	162	6 AX916981	AX916981 Sequence

# ALIGNMENTS

RESULT 1

AF097432	AF097432	2600 bp	mRNA	linear	PRI 22-NOV-2000
LOCUS	AF097432	2600 bp	mRNA	linear	PRI 22-NOV-2000
DEFINITION	Homo sapiens GROS1-L protein mRNA, complete cds.				
ACCESSION	AF097432				
VERSION	AF097432.1	GI:11127637			
KEYWORDS					
SOURCE	Homo sapiens (human)				
ORGANISM	Homo sapiens				
REFERENCE	1 (bases 1 to 2600)				
AUTHORS	Kaul, S.C., Sugihara, T., Yoshida, A., Nomura, H. and Wadhwa, R.				
TITLE	Gros1, a potential growth suppressor on chromosome 1: its identity to basement membrane-associated proteoglycan, leprecan				
JOURNAL	Oncogene 19 (32), 3576-3583 (2000)				
MEDLINE	20406537				
PUBMED	10951563				
REFERENCE	2 (bases 1 to 2600)				
AUTHORS	Wadhwa, R., Sugihara, T. and Kaul, S.				
TITLE	Direct Submission				
JOURNAL	Submitted (06-OCT-1998) Group 4, Chugai Research Institute of Molecular medicine, 153-2 Nagai, Niihari, Ibaraki 300-41, Japan				
FEATURES	Location/Qualifiers				
source	1..2600				
	/organism="Homo sapiens"				
	/mol_type="mRNA"				
	/db_xref="taxon:9606"				
	/chromosome="1"				
	/tissue_type="testis"				
	52..2262				
	/note="member of protein B family"				
	/codon_start=1				
	/product="GROS1-L protein"				
	/protein_id="BAG31019.1"				
	/db_xref="GI:11127638"				
	/translation="MAVRAKLLTLAVVAASQAQVESEAGKGMVTPDLLPAEGTA AYARDGFWGLSERRAALRLRCRTQCAADFWEIDPDWSPQAQAG ALRDSFGGLRRAACLRCLGPPAAHLSSEMELEFRKRSYNTLYQAVFKINKLE KAVAAHTFVGNPEHMEMQNDLYQYTMGSKVEADPKLETPHMOERFLGVRLYSE EPOEAAPHLEAALQEVFAVEECRLCEGPDYDGVNLYENADLFOAITDHIQVIL NCKQCVTELASHRSRKPDEPLPSHYNLYQFAYNIGNYTAGSCAKTYLLFPFND EVMQNLYAAMLEBHTSISGPRSAKRYRQSLKLELFFAIDVFGIPFVDPDS WTPVEVLPKRLQKQSERETAVRSOIEIGNLKMIEITLVEEKTKESLDVSRLTRGG PLLYEGISLTNKLNGYQRYVMDGVI SDHEQELQRLTNVAATSDGVRGTSPT PNEKFGYTVFKALKQEGKVPLOSAHLYNIVTEKVRIMESYFRLDTPLYFSYHL VCRTAIEVQAEKDDSHPVNDICILNAETLVCKVEPPAYTPRDSAILYLNGDFDG GNFYETLDAKTVTAEVQPCGGRAGVFGSSGTENPHGVKAVTRGRCRAIALWFLDPRH SERDEVDADLDVLMFLSPFEENDLSQEPDLAQOQPEPPAQESLSGESKPKDEL"				
ORIGIN					
Alignment Scores:					
Pred. No.:	0	Length:	2600		
Score:	736.00	Matches:	736		
Percent Similarity:	100.00%	Conservative:	0		
Best Local Similarity:	100.00%	Mismatches:	0		
Query Match:	100.00%	Indels:	0		
DB:	9	Gaps:	0		
US-10-045-815-4 (1-736) x AF097432 (1-2600)					
Qy	1	MetAlaValArgAlaLeuLysLeuLeuThrThrLeuLeuAlaValAlaAlaSer	20		
Db	52	ATGGCGGTACCGCGTTGAAGCTGCTGACCACTGCTGGCTGCGGCGGCTGCTCC	111		
Qy	21	GlnAlaGluValGluSerGluAlaGlyTrpGlyMetValThrProAspLeuLeuPheAla	40		
Db	112	CAAGCCGAGTCCGAGTCCGAGCAGGATGGGCATGGCGCTGATCTGCTCTTCGCC	171		
Qy	41	GluGlyThrAlaAlaValAlaArgGlyValAspTrpProGlyValValLeuSerMetGluArg	60		
Db	172	GAGGGACCGGCTTACGCGCGGGGACTGGCCCGGGTGGTCTTGACATGGAGACGG	231		
Qy	61	AlaLeuArgSerArgAlaAlaLeuArgAlaLeuArgLeuArgCysArgThrGlnCysAla	80		
Db	232	GGGCTCGCTCCCGGCGAGCCCTCCGCGCCCTTCGCTCGCTCGCCACCCAGTGTGCC	291		
Qy	81	AlaAspPheProTrpGluLeuAspProAspTrpSerProSerProAlaGlnAlaSerGly	100		
Db	292	GCCGACTTCCCGTGGAGCTGGACCCGACTGGTCCCCCAGCCCGCCGCGCTCGGCG	351		
Qy	101	AlaGlyAlaLeuArgAspLeuSerPhePheGlyGlyLeuLeuArgArgAlaAlaCysLeu	120		
Db	352	GCGGCGCCCTCGGCGACCTGAGCTTCTCGGGGCTTCTCGCTCGGCTGCTCGCTG	411		
Qy	121	ArgArgCysLeuGlyProProAlaAlaHisSerLeuSerGluGluMetGluLeuPhe	140		
Db	412	CGCCGCTGCTCGGCGCGCGCCGCCCTCCCTCAGCGAAGAGATGGAGCTGGAGTTC	471		
Qy	141	ArgLysArgSerProTrpAsnTrpLeuGlnValAlaValPheLysIleAsnLysLeuGlu	160		
Db	472	CGCAAGCGGAGCCCTACAACTACCTGAGGTGGCTACTTCAAGATCAACAAGTTGGAG	531		
Qy	161	LysAlaValAlaAlaAlaHisThrPhePheValGlyAsnProGluHisMetGluMetGln	180		
Db	532	AAAGCTGTGCTGCAGCACACACCTTCTTCGGGGCAATCTCTGACACATGGAATGCG	591		
Qy	181	GlnAsnLeuAspTrpTrpGlnThrMetSerGlyValLysGluAlaAspPheLysAspLeu	200		
Db	592	CAGAACCTAGACTATTACCAACCATGCTGGAGTGAAGGAGCGGCTTCAAGGATCTT	651		
Qy	201	GluThrGlnProHisMetGlnGluPheArgLeuGlyValArgLeuTrpSerGluGluGln	220		
Db	652	GAGACTCAACCCCATATCAAGATTTTCGACTGGGAGTGGCACTTACTCTAGAGGAACAG	711		
Qy	221	ProGlnGluAlaValProHisLeuGluAlaLeuGlnGluTrpPheValAlaValTrpGlu	240		
Db	712	CCACAGGAAGCTGTGCCCCACCTAGAGGGGCGCTGCAAGATATCTTGTGGCTATGAG	771		
Qy	241	GluCysArgAlaLeuCysGluGlyProTrpAspTrpAspGlyTrpAsnTrpLeuGluTrp	260		
Db	772	GAGTCCGCTCGCTCGGAAAGGCGCTTACTGACTACGATGGCTACAACACTCTGAGTAC	831		
Qy	261	AsnAlaAspLeuPheGlnAlaValThrAspHisTrpIleGlnValLeuAsnCysLysGln	280		
Db	832	AAAGCTGACCTTTCAGGCCATCACAGATCATTACATCCAGTCTCTCACTGTAAGCAG	891		
Qy	281	AsnCysValThrGluLeuAlaSerHisProSerArgGlyLysProPheGluAspPheLeu	300		
Db	892	AACTGTGTACGAGCTTGTCTCCACCAAGTCGAGAGAACCCCTTTGAAGACTTCTCTC	951		
Qy	301	ProSerHisTrpAsnTrpLeuGlnPheAlaValTrpAsnIleGlyAsnTrpThrGlnAla	320		
Db	952	CCATCGCATTAATATCTGAGTTTGGCTTACTATAACATTTGGAAATATACCAAGCT	1011		
Qy	321	GlyGluCysAlaLysThrTrpLeuLeuPhePheProAsnAspGluValMetAsnGlnAsn	340		
Db	1012	GGTGAATGTCCACAGACCTATCTTCTTCTTCCCAATGACGAGGTGATGAACCAAAAT	1071		
Qy	341	LeuAlaValTrpAlaAlaMetLeuGlyGluHisThrArgSerIleGlyProArgGlu	360		
Db	1072	TTGGCTTATTATGACGCTATGCTTGGAGAGAAACACACAGATCCATCGGCCCGCGTGG	1131		
Qy	361	SerAlaLysGluTrpArgGlnArgSerLeuLeuGluLysGluLeuLeuPhePheAlaTrp	380		
Db	1132	AGTGCAAGGAGTACCGACAGCGAAGCTACTGGAAAAAGAACTGCTTTTCTTCGCTTAT	1191		
Qy	381	AspValPheGlyIleProPheValAspProAspSerTrpThrProGluGluValIlePro	400		
Db	1192	GATGTTTTTGGAAATTCCTTTTGTGATCCGGAATTCATGGACTCCAGAGAGATGATCCC	1251		
Qy	401	LysArgLeuGluLysGlnLysSerGluArgGluThrAlaValArgIleSerGlnGlu	420		
Db	1252	AGAGATTGCAAGAGAAACAGAGTCAAGACGGGAACAGCGTACGCACTCCCGAGGAG	1311		
Qy	421	IleGlyAsnLeuMetLysGluIleGluThrLeuValGluGluLysThrLysGluSerLeu	440		

Db 1312 ATTGGGAACCTTATGAAGAAATCGAGAACCTTTGTGGAAGAGAACCAAGGAGTCACTG 1371

Qy 441 AspValSerArgLeuThrArgGluGlyGlyProLeuLeuTyrGluGlyLeuSerLeuThr 460

Db 1372 GATGTGAGCAGATGATCCCGGGAAGGTGGCCCTCTGTGTATGAAGCATCAGTCTCACC 1431

Qy 461 MetAsnSerLysLeuLeuAsnGlyTyrGlnArgValMetAspGlyValLleSerAsp 480

Db 1432 ATGAACCTCCAAATCTCTGAATGTTACCAAGCGGTGTGATGACGCGCTAATCTCTGAC 1491

Qy 481 HisGluCysGlnGluLeuGlnArgLeuThrAsnValAlaThrSerGlyAspGlyTyr 500

Db 1492 CAGAGTGTGAGGAGTGCAGAGACTGACCAATGTGGCAGCAACCTCAGAGATGGCTAC 1551

Qy 501 ArgGlyGlnThrSerProHisThrProAsnGluLysPheTyrGlyValThrValPheLys 520

Db 1552 CGGGGTGAGACCTCCACATATCTCCCAATGAAAGTTCTATGCTGCTACTGCTTCAA 1611

Qy 521 AlaLeuLysLeuGlyGlnGluGlyLysValProLeuGlnSerAlaHisLeuTyrTyrAsn 540

Db 1612 GCCCTCAAGCTGGGGCAAGCAAGTTCCTCTGAGAGTGCCACCTGTACTACAAC 1671

Qy 541 ValThrLysValArgArgIleMetGluSerTyrPheArgLeuAspThrProLeuTyr 560

Db 1672 GTGACGGAGAAAGTGGGGCCATCATGAGTCTCTACTTCGCTGTGATAGCCCTCTAC 1731

Qy 561 PheSerTyrSerHisLeuValCysArgThrAlaIleGluGluValGlnAlaGluArgLys 580

Db 1732 TTTTCTCTACTCTCATCTGTGTGCGCATCTGCCATCGAAGAGTCCAGGAGAGAG 1791

Qy 581 AspAspSerHisProValHisValAspAsnCysIleLeuAsnAlaGluThrLeuValCys 600

Db 1792 GATGATAGTATCATCAGTCCACGTGGACACTGCTATCTGATGCCGAGACCTCTGCTGT 1851

Qy 601 ValLysGluProProAlaTyrThrPheArgAspTyrSerAlaIleLeuTyrLeuAsnGly 620

Db 1852 GTCAAGAGAGCCCGAGCTACACTTCGCGGACTACAGCGCCATCTTTACCTAAATGG 1911

Qy 621 AspPheAspGlyGlyAsnPheTyrPheThrGluLeuAspAlaLysThrValThrAlaGlu 640

Db 1912 GACTTCGATCGCGAAACTTTTATTTTCACTGAATCGATGCCAAGACCGTGACGGCAG 1971

Qy 641 ValGlnProGlnCysGlyArgAlaValGlyPheSerSerGlyThrGluAsnProHisGly 660

Db 1972 GTGACGCTCAGTGTGAAGAGCGTGGGATCTCTTCAGGCATGAAACCCACATGGA 2031

Qy 661 ValLysAlaValThrArgGlyGlnArgCysAlaIleAlaLeuThrPheThrLeuAspPro 680

Db 2032 GTGAAGGCTGTCAACAGGGGCGAGCGCTGTGCCATCGCCCTGTGGTTCACCCCTG 2091

Qy 681 ArgHisSerGluArgAspArgValGlnAlaAspLeuValLysMetLeuPheSerPro 700

Db 2092 CGACACAGCAGCGGGACAGGGTGCAGGCAGATGACCTGTGAAGATGCTTTCAGCCCA 2151

Qy 701 GluGluMetAspLeuSerGlnGluProLeuAspAlaGlnGlnGlyProProGluPro 720

Db 2152 GAAGAGATGACCTCTCCAGGAGCAGCCCTGATGCCAGAGAGGCCCCCGAACCT 2211

Qy 721 AlaGlnGluSerLeuSerGlySerGluSerLysProLysAspGluLeu 736

Db 2212 GCACAGAGTCTCTCTCAGGAGTGAATCGAAGCCCAAGGATGAGCTA 2259

RESULT 2

AX136191

LOCUS AX136191 2563 bp DNA linear PAT 30-MAY-2001

DEFINITION Sequence 113 from Patent EP1067182.

ACCESSION AX136191

VERSION AX136191.1 GI:14272599

KEYWORDS

SOURCE Homo sapiens (human)

ORGANISM Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE 1

AUTHORS Ota,T., Isogai,T., Nishikawa,T., Kawai,Y., Sugiyama,T. and Hayashi,K.

TITLE Secretory protein or membrane protein

JOURNAL Patent: EP 1067182-A 113 10-JAN-2001;

FEATURES

source Helix Research Institute (JP)

Location/Qualifiers

1..2563

/organism="Homo sapiens"

/mol\_type="unassigned DNA"

/db\_xref="taxon:9606"

31..2241

/note="unnamed protein product"

/codon\_start=1

/protein\_id="CAC39738.1"

/db\_xref="GI:14272600"

/db\_xref="REMBL: CAC39738"

/translation="MAVRALKLTLTLAAVAAASQAEVESEAGWMTPTDILFAEGTA

AVARGDPEGVLSMERALSRALRLRCRTCAADFPWELDPDMSPPSAQASGAA

ALRLSPFGGLLRRAACLRCLGPPAAHSISEMELFRKSPYNYLQVAVFKINLE

KAVAAAHFTFVGNPHEMEOQLDYIQTMSGVKEADKDLQETQPMQSEFRLGVLYSE

EQPEAVPHEALQEVFVAYEECALCEGYDYDGYNYLEYNADLFQALTDHIXQL

NCKQCVTELAHSPSRKPFEDFLPSHNYLQFAYINIGNYQVCAKTYLLFFPND

EWNONLAYAAMLGESHTRSIGPRESAKYRQSLLEKLLFFAYDVGTFPFVDPDS

WTPBEVTPKLOEKOKSERFAVRISQIEIGNKEIETLVEKTKESLDVSRLTRGG

PLLVEGSLTWNSKLNGSORVMDGVTSDEHCELOELTNVAATSGDVGRTGSPHT

ENEXFYGVTFKALKLQEGCKVPLOSAHLVYNTVEKVRIMESYERLDTPLYFSYSL

VCRTALBEVQERKDDSHPHVDNCINLAETLVCVKPPATFRDYSAILLVNGDFDG

GNFYTEDAAKTVAEVOPOCGRAVFGFSFSTENPHGVKAVTRGRCALMFTLDPRH

SERDRVQADDLVKMLFSPPEMDLSQEQFLDAQQGPPEPPEAQESLSGESKPKDEL"

ORIGIN

Alignment Scores:

Pred. No.: 0 Length: 2563

Score: 433.00 Matches: 733

Percent Similarity: 99.19% Conservative: 0

Best Local Similarity: 99.19% Mismatches: 3

Query Match: 58.83% Indels: 6

DB: 6 Gaps: 0

US-10-045-815-4 (1-736) x AX136191 (1-2563)

Qy 1 MetAlaValArgAlaLeuLysLeuLeuThrThrLeuLeuAlaValAlaAlaSer 20

Db 31 ATGGCGGTACGCGCTTGAAGCTGTGACCACTGCTGGCTGTGTCGCGCGCTGCCCTCC 90

Qy 21 GlnAlaGluValGluSerGluAlaGlyTyrGlyMetValThrProAspLeuLeuPheAla 40

Db 91 CAAGCCGAGGTGAGTCCGAGCAGATGGGCGATGGTGACCGCTGATCTGCTCTTCGCC 150

Qy 41 GluGlyThrAlaAlaTyrAlaArgGlyAspTyrProGlyValValLeuSerMetGluArg 60

Db 151 GAGGGACCGCAGCTACGCGCGGGGACTGGCCCGGGTGGTCTCAGCATGGAACGG 210

Qy 61 AlaLeuArgSerArgAlaAlaLeuArgAlaLeuArgLeuArgCysArgThrGlnCysAla 80

Db 211 GCGCTCGCTCCGCGGAGCCCTCCGCGCCCTTCGCCCTGCCCTGCCAGCGTGGCC 270

Qy 81 AlaAspPheProTyrGluLeuAspProAspTyrProSerProAlaGlnAlaSerGly 100

Db 271 GCCGACTTCCGTGGAGACTGGACCCCGACTGGTCCCCCAGCCCGCCCGCTCGGGC 330

Qy 101 AlaGly-AlaLeuArgAspLeuSerPheGlyGlyLeuLeuArgArgAlaAlaCysLe 120

Db 331 GCCCGC-CGCGCTCGCGACCTGAGCTTCTTCGGGGCGCTTCTCGCTCGCGCTCGCT 389

Qy 120 uArgArgCysLeuGlyProProAlaAlaHisSerLeuSerGluGluMetGluLeuGluPh 140

Db 390 GCGCGCTGCTCGGGCGCGCGCGCCACTTCGCTCAGCGAAGAGATGGAGCTGGAGTT 449

Qy 140 eArgLysArgSerProTyrAsnTyrLeuGlnValAlaTyrPheLysIleAsnLysLeuGl 160

Db 450 CCGCAAGCGGAGCCCTTACAACTACCTGAGGTGCGCTACTTCAAGATCAACAGTTGGA 509

160 uLysAlaValAlaAlaAlaHisThrPhePheValGlyAsnProGluHisMetGluMetG1 180  
 180 nGlnAsnLeuAspTyrGlnThrMetSerGlyValLysGluAlaAspPheLysAspLe 200  
 570 GCAGAACCTAGACTATTACCAAAACCATCTCTGGAAGTGAAGAGGCGGACCTTCAAGGATCT 629  
 200 uGluThrGlnProHisMetGlnGluPheArgLeuGlyValArgLeuTyrSerGluGluG1 220  
 630 TGAGACTCAACCCCATATGCAGAAATTTGAGTGGAGTGGACTTACTCAGAGGAACA 689  
 220 nProGlnGluAlaValProHisLeuGluAlaAlaLeuGlnGluTyrPheValAlaTyrG1 240  
 690 GCCACAGGAAGCTGTGCCCCACCTAGAGGCGGCGCTGCAAGAAATCTTTGTGGCTATGA 749  
 240 uGluCysArgAlaLeuCysGluGlyProTyrAspTyrAspGlyTyrAsnTyrLeuGluTyr 260  
 750 GGAGTGGCGTGCCTCTCGGAAGGCGCTATGACTAGATGGCTACAACTACCTTTGAGTA 809  
 260 rAsnAlaAspLeuPheGlnAlaIleThrAspHisTyrIleGlnValLeuAsnCysLysG1 280  
 810 CAACGCTGACCTCTCCAGGCGCATCACAGATCATTTACATCCAGTCTCAACTGTAAAGCA 869  
 280 nAsnCysValThrGluLeuAlaSerHisProSerArgGluLysProPheGluAspPheLe 300  
 870 GAACCTGTGTACGGAGCTTGCTTCCCAACCCCAAGTCGAGAGAGCCCTTTTGAAGACTTCCT 929  
 300 uProSerHisTyrAsnTyrLeuGlnPheAlaTyrTyrAsnIleGlyAsnTyrThrGlnAl 320  
 930 CCCATCGCATTAATTAATCTCGAGTTGCCATTAATCAATTCGGAAATATACACAGGC 989  
 320 aGlyCysAlaLysThrTyrLeuLeuPhePheProAsnAspGluValMetAsnGlnA 340  
 990 TGT-TGAATGTGCCAAGACCTTATCTTCTTCTTCCCAATGACAGGCTGATCAACCAAA 1048  
 1049 ATTTGGCTTATTATGAGCTATGCTTGGAGAAACACACAGATCCATCGGCCCGCTG 1108  
 360 l uSerAlaLysGluTyrArgGlnArgSerLeuLeuGluLysGluLeuPhePheAlaT 380  
 1109 AGATGCCAAGAGTACCGACAGGAAGCTTACTTGGAAAGAAAGAACTCTTTTCTTGGCTT 1168  
 380 yrAspValPheGlyIleProPheValAspProAspSerThrProGluGluValIleP 400  
 1169 ATGATGTTTTTGGAAATTCCTTTTGTGATCGGATTCATGATCCAGAGAAAGTGAATC 1228  
 400 r oLysArgLeuGlnGluLysGlnLysSerGluArgGluThrAlaValArgIleSerGlnG 420  
 1229 CCAAGAGATTGCAGAGAAACAGAGTACAGCGGAAACAGCGTACGATCTCCAGG 1288  
 420 l uIleGlyAsnLeuMetLysGluIleGluThrLeuValGluLysThrLysGluSerL 440  
 1289 AGATTGGGAACCTTATGAAGAAATCAGACCTTGTGGAAGAGAAACCAAGAGTCCAC 1348  
 440 euAspValSerArgLeuThrArgGluGlyProLeuLeuTyrGluGlyIleSerLeuT 460  
 1349 TGGATGTGAGAGACTGACCGGGAAGTGGCCCCCTGCTGATGAAGCATCAGTCTCA 1408  
 460 hrMetAsnSerLysLeuLeuAsnGlyTyr-GlnArgValValMetAspGlyValIleSer 479  
 1409 CCATGAACCTCAAACTCCTGAAATGGTTTC-CCAGCGGTGGTGTGAGAGCGCGTAAATCTCT 1467  
 480 AspHisGlyCysGlnGluLeuGlnArgLeuThrAsnValAlaAlaThrSerGlyAspGly 499  
 1468 GACCACGAGTGTACGAGCTGCAGAGACTGACCAATGTGCAGCAACCTCAGAGATGGC 1527  
 500 TyrArgGlyGlnThrSerProHisThrProAsnGluLysPheTyrGlyValThrValPhe 519  
 1528 TACCGGGGTGAGACCTCCCAACATCTCCCAATGAAAAGTTCTATGGTGTCACTGTCTTC 1587

520 LysAlaLeuLysLeuGlyGlnGluLysValProLeuGlnSerAlaHisLeuTyrTyr 539  
 1588 AAAGCCCTCAAGCTGGGCAAGAGCAAGTTCCTCTGCAGAGTGGCCACCTGTACTAC 1647  
 540 AsnValThrGluLysValArgIleMetGluSerTyrPheArgLeuAspThrProLeu 559  
 1648 AACGTGACGAGGAAGTGGCGGCATCATGGAGTCTTCTTCCGCTGGATACGCCCTC 1707  
 560 TyrPheSerTyrSerHisLeuValCysArgThrAlaIleGluGluValGluAlaGluArg 579  
 1708 TACTTTTCTTACTCTCACTGCTGCGCACTGCATCGAAGAGTCCAGCAGAGAGG 1767  
 580 LysAspAspSerHisProValHisValAspAsnCysIleLeuAsnAlaGluThrLeuVal 599  
 1768 AAGGATGATAGTCAATCCAGTCCAGTGGACAACTCATCTGAATGCCAGACCCCTCGT 1827  
 600 CysValLysGluProProAlaTyrThrPheArgAspTyrSerAlaIleLeuTyrLeuAsn 619  
 1828 TGTGTCAAGAGACCCCGCCAGCTTACCTTCCCGCACTACAGCCCATCTTTTACCCTAAT 1887  
 620 GlyAspPheAspGlyGlyAsnPheTyrPheThrGluLeuAspAlaLysThrValThrAla 639  
 1888 GGGGACTTCGATGGCGGAAACTTTTATTTCTGAATCGATGCCAAGACCGTGACGCA 1947  
 640 GluValGlnProGlnCysGlyArgAlaValGlyPheSerSerGlyThrGluAsnProHis 659  
 1948 GAGGTGACGCTCAGTGTGAAGAGCCGTGGATTTCTTTCAGGCACCTGAAGACCCCAT 2007  
 660 GlyValLysAlaValThrArgGlyGlnArgCysAlaIleAlaLeuTyrPheThrLeuAsp 679  
 2008 GGAAGTGAAGCTGTACCCAGGGGGAGCGCTGTGCCATCGCCCTGTGTTTCCCTTGGAC 2067  
 680 ProArgHisSerGluArgAspArgValGlnAlaAspAspLeuValLysMetLeuPheSer 699  
 2068 CCTGCACACAGCGAGCGGACAGGCTGTCAGGAGACCTGCTGGTGAAGATGCTCTTTCAGC 2127  
 700 ProGluGluMetAspLeuSerGlnGluGlnProLeuAspAlaGlnGlnGlyProGlu 719  
 2128 CCAGAAGAGATGAGCTCTCCAGGAGAGCCCTGATGCCAGAGGCGGCCCGCCGAA 2187  
 720 ProAlaGlnGluSerLeuSerGlySerGluSerLysProLysAspGluLeu 736  
 2188 CCTGCACAAGAGTCTCTCTCAGGAGTGAATCGAAGCCCAAGAGTGAAGTGA 2238  
 RESULT 3  
 BDI23548  
 LOCUS  
 DEFINITION  
 SECRETORY protein or membrane protein.  
 BDI23548  
 ACCESSION  
 BDI23548.1 GI:23218493  
 VERSION  
 JP 2002017376-A/57.  
 KEYWORDS  
 Homo sapiens (human)  
 SOURCE  
 ORGANISM  
 Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 REFERENCE  
 1 (bases 1 to 2563)  
 AUTHORS  
 Ota,T., Isogai,T., Nishikawa,T., Kawai,Y., Sugiyama,T. and  
 Hayashi,K.  
 TITLE  
 Secretory protein or membrane protein  
 JOURNAL  
 Patent: JP 2002017376-A 57 22-JAN-2002;  
 HELIX RESEARCH INSTITUTE  
 COMMENT  
 OS Homo sapiens (human)  
 PN JP 2002017376-A/57  
 PD 22-JAN-2002  
 PF 07-JUL-2000 JP 2000253173  
 PI TOSHIO OTA,TAKAO ISOGAI,TETSUO NISHIKAWA,YURI KAWAI,TOMOYASU  
 PI SUGIYAMA,  
 PI KOJI HAYASHI  
 PC  
 C12N15/09, C07K14/47, C07K16/18, C12N1/15, C12N1/19, C12N1/21, C12N5/ PC  
 10,  
 PC C12P21/02, C12Q1/68//C12P21/08, C12N15/00, C12N5/00 CC  
 Secretory protein or membrane protein



PH	Key	Location/Qualifiers	Location/Qualifiers	Location/Qualifiers
FT	CDS			
FEATURES				
source				
		1. .2563	/organism="Homo sapiens"	
		/mol_type="genomic DNA"		
		/db_xref="taxon:9606"		
ORIGIN				
Alignment Scores:				
Pred. No.:	0	Length:	2563	
Score:	433.00	Matches:	733	
Percent Similarity:	99.19%	Conservative:	0	
Best Local Similarity:	99.19%	Mismatches:	3	
Query Match:	58.83%	Indels:	6	
DB:	6	Gaps:	0	
US-10-045-815-4 (1-736) x BD123548 (1-2563)				
QY	1	MetAlaValArgAlaLeuLysLeuLeuThrThrLeuLeuAlaValValAlaAlaLasar	20	
DB	31	ATGGCGGTACGCGGTTGAAGCTGTGACACACTGTGGCTCTCGTGGCGGCTGCC	90	
QY	21	GlAlaGluValGluSerGluAlaGlyTTPGlyMetValThrProAspLeuLeuPheAla	40	
DB	91	CAAGCCGAGGTCGAGTCGAGCGGAGGATGGGGCATGGTGACGCTGATCTGCTCTCGCC	150	
QY	41	GluGlyThrAlaAlaValAlaArgGlyAspTTPProGlyValValLeuSerMetGluArg	60	
DB	151	GAGGGACCCGACCTACGCGCGCGGGGACTGGCCCGGGGTGGTCTGAGCATGGAA	210	
QY	61	AlaLeuArgSerArgAlaAlaLeuArgAlaLeuArgLeuArgCysArgThrGlnCysAla	80	
DB	211	CGCGTGGCTCCCGGGGAGCCCTCGCGGCTTCGCGGCTTCGCGGCTCGCGGCTCGCC	270	
QY	81	AlaAspPheProTTPGluLeuAspProAspTTPSerProAlaGlnAlaSerGly	100	
DB	271	GCGGACTTCCCGTGGAGCTGGACCCCGACTGGTCCCCAGCCCGCGGCTCGGGC	330	
QY	101	AlaGly-AlaLeuArgAspLeuSerPhePheGlyGlyLeuLeuArgAlaAlaCysLe	120	
DB	331	GGCGC-CGCGCTCGCGGACCTGAGCTTCTTCGGGGGCTTCGCGGCTCGCGGCT	389	
QY	120	uArgArgCysLeuGlyProProAlaAlaHisSerLeuSerGluGluMetGluLeuGluPh	140	
DB	390	GCGCGCTGCTCGCGGCGCGCGCCCTCGCTCGCTCAGCAAGAGATGGAGCTGGAGT	449	
QY	140	eArgLysArgSerProTTPAsnTTPLeuGlnValAlaValPheLysIleAsnLysLeuG	160	
DB	450	CCGCAAGCGGAGCCCTACACTACTCTGAGGTGCGCTCTTCAAGATCAACAGTTGGA	509	
QY	160	uLysAlaValAlaAlaHisThrPhePheValGlyAsnProGluHisMetGluMetG	180	
DB	510	GMAAGCTGTTGTCGACGACACACCTTCTTCGTTGGGCAATCTCTGAGCACATGGA	569	
QY	180	nGlnAsnLeuAspTTPThrGlnThrMetSerGlyValLysGluAlaAspPheLysAsp	200	
DB	570	GCAGAACTAGACTATTACCAACCATGTCTGAGTGAAGAGCGCCACTTCAAGATCT	629	
QY	200	uGluThrGlnProHisMetGlnGluPheArgLeuGlyValArgLeuTyrSerGluGlu	220	
DB	630	TGAGACTCAACCCCATATGCAAGATTTTCGACTGGGAGTGGGCTCTACTCAGAGGA	689	
QY	220	nProGlnGluAlaValProHisLeuGluAlaAlaLeuGlnGluTTPPheValAlaTyrG	240	
DB	690	GCCACAGAAAGCTGTGCCCCACCTAGAGCGCGGCTGCAAGAATACTTTGTGGCC	749	
QY	240	uGluCysArgAlaLeuCysGluGlyProTTPAspTTPAspGlyTTPAsnTTPLeuGlu	260	
DB	750	GGAGTGGCGGCTCTTCGAGAGGCGGCTTACTACTAGATGGCTACACTACCTTGA	809	
QY	260	rAsnAlaAspLeuPheGlnAlaIleThrAspHisTTPLeGlnValLeuAsnCysLysG	280	

DB	810	CAACGCTGACCTCTTCCAGGCGCATCACAGATCATTTACATCCAGGTCCTCAACTGTAGCA	869	
QY	280	nAsnCysValThrGluLeuAlaSerHisProSerArgGluLysProPheGluAspPheLe	300	
DB	870	GAACTGTGTACGAGGCTTGTCTCCACCCAGTCGAGAGGCGCTTTGAAGACTTCTCT	929	
QY	300	uProSerHisTyrAsnTTPLeuGlnPheAlaTyrTyrAsnIleGlyAsnTTPThrGlnAl	320	
DB	930	CCCATCGCATTAATATCTGCACTTTCGCTACTATAACATTGGGAATTATACACAGC	989	
QY	320	aGly-GluCysAlaLysThrTyrLeuLeuPhePheProAsnAspGluValMetAsnGlnA	340	
DB	990	TGT-TGAATGTGCCAAGACCTATCTCTCTTCTTCCCCCAATGACGAGGTGATGAACCA	1048	
QY	340	snLeuAlaTyrTyrAlaAlaMetLeuGlyGluGluHisThrArgSerIleGlyProArgG	360	
DB	1049	ATTGGCCCTATTATGACAGCTATGCTTGAGAGAACAACACAGATCCATCGCCCCCGT	1108	
QY	360	luSerAlaLysGluTyrArgGlnArgSerLeuLeuGluLysGluLeuLeuPheAlaI	380	
DB	1109	AGAGTGCCCAAGAGTACCGACAGCGAAGCTACTGGAAAAAGAACTGCTTTCTTCGCT	1168	
QY	380	yrAspValPheGlyIleProPheValAspProAspSerTTPThrProGluGluValIleP	400	
DB	1169	ATGATGTTTTTGGAAATCCCTTTTGGATCCGGATTCTATGACTCCAGAGAGTGTTC	1228	
QY	400	roLysArgLeuGlnGluLysGlnLysSerGluArgGluThrAlaValArgIleSerGlnG	420	
DB	1229	CCAAGAGATTGCAAGAGAAACAGAAAGTCAGAAAGCCCTGTGGAAGAGAGAACCAAG	1288	
QY	420	luIleGlyAsnLeuMetLysGluIleGluThrLeuValGluGluLysThrLysGluSerL	440	
DB	1289	AGATTGGGAACCTTATGAAGAAATCGAGACCCCTGTGGAAGAGAGAACCAAGAGT	1348	
QY	440	euAspValSerArgLeuThrArgGluGlyGlyProLeuLeuTyrGluGlyIleSerLeu	460	
DB	1349	TGGATGTGACGAGACTGACCCGGGAGGTGGCCCTCTGCTGTATGAGGATCATCTCA	1408	
QY	460	hrMetAsnSerLysLeuLeuAsnGlyTyr-GlnArgValValMetAspGlyValIleSer	479	
DB	1409	CCATGAACCTCCAACTCCTGAATGCTTC-CCAGCGGCTGGTGTGATGGAGCGGCTA	1467	
QY	480	AspHisGluCysGlnGluLeuGlnArgLeuThrAsnValAlaAlaThrSerGlyAspGly	499	
DB	1469	GACCAAGAGTGTGAGAGCTGCAGAGACTGCACCAATGTGGCAGCAACCTCAGGAGT	1527	
QY	500	TyrArgGlyGlnThrSerProHisThrProAsnGluLysPheTyrGlyValThrValPhe	519	
DB	1528	TACCGGGTTCAGACTCCCTCCCATCTCCCAATGAAAGTTCTATGTGTCTCTCTTC	1587	
QY	520	LysAlaLeuLysLeuGlyGlnGluLysValProLeuGlnSerAlaHisLeuTyrTyr	539	
DB	1588	AAAGCCCTCAAGCTGGGCAAGAGCAAGTTCCTCTGCAGAGTGCCCATCTGTACTAC	1647	
QY	540	AsnValThrGluLysValArgArgIleMetGluSerTyrPheArgLeuAspThrProLeu	559	
DB	1648	AACTGACGAGAAAGTGGCGGCATCATGGAGTCTCTACTCCGCTGGATACGCCCTC	1707	
QY	560	TyrPheSerTyrSerHisLeuValCysArgThrAlaIleGluGluValGlnIleGluArg	579	
DB	1708	TACTTTTCT	1767	
QY	580	LysAspAspSerHisProValHisValAspAsnCysIleLeuAsnAlaGluThrLeuVal	599	
DB	1768	AAGATGATAGTCATCCAGTCCAGTGCACACTGCATCTCTGAAATGCCGAGACCTCTG	1827	
QY	600	CysValLysGluProProAlaTyrThrPheArgAspTyrSerAlaIleLeuTyrLeuAsn	619	
DB	1828	TGTGTCAAAGAGCCCGAGCTTACACTTCCGCGACTACAGCGCATCTCTTTTACCTAA	1887	
QY	620	GlyAspPheAspGlyGlyAsnPheTyrPheThrGluLeuAspAlaLysThrValThrAla	639	
DB	1888	GGGAGCTTCGATGGCGGAAACTTTTATTTCATCTGAATGATGCCAAGACCTGAGGCA	1947	

QY 640 GluValGlnProGlnCysGlyArgAlaValGlyPheSerSerGlyThrGluAsnProHis 659  
 DB 1948 GAGTGGACGCTCAGTGTGGAGAGCGCGTGGGATCTCTTTCAGGACCTGAAACCCACAT 2007  
 QY 660 GlyValLysAlaValThrArgGlyGlnArgCysAlaIleAlaLeuTTPheThrLeuAsp 679  
 DB 2008 GGAGTGAAGGCTGTTCACACAGGGGCGAGCGCTGTGCATCGCCCTGTGTTACCCCTGCAC 2067  
 QY 680 ProArgHisSerGluArgAspArgValGlnAlaAspAspLeuValLysMetLeuPheSer 699  
 DB 2068 CCTCGACACAGCGAGCGGACAGGGTGCAGGACAGTACCTGTGTGAAGATGCTCTTCAGC 2127  
 QY 700 ProGluGluMetAspLeuSerGlnGlnProLeuAspAlaGlnGlnGlyProProGlu 719  
 DB 2128 CCAGAGAGATGGACCTCTCCAGAGAGAGCGCCCTGTGATGCCAGAGGCGCCCGCGAA 2187  
 QY 720 ProAlaGlnGluSerLeuSerGlySerGluSerLysProLysAspGluLeu 736  
 DB 2188 CCTGCACAAGATCTCTCTCAGGAGTGAATCGAAGCCCAAGGATGAGCTA 2238  
 RESULT 4  
 AK075418 2563 bp mRNA linear PRI 03-SEP-2002  
 LOCUS Homo sapiens CDNA PSEC0109 fis, clone NT2RP2005671, moderately  
 DEFINITION similar to Rattus norvegicus leprecan (leprel) mRNA.  
 ACCESSION AK075418  
 VERSION AK075418.1 GI:22761493  
 KEYWORDS oligo capping; fis (full insert sequence).  
 SOURCE Homo sapiens (human)  
 ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 1  
 Ota, T., Nishikawa, T., Suzuki, Y., Kawai-Hio, Y., Hayashi, K.,  
 Ishii, S., Saito, K., Yamamoto, J., Wakamatsu, A., Nagai, T.,  
 Nakamura, Y., Nagahara, K., Sugano, S. and Isogai, T.  
 HRI human cDNA sequencing project  
 Unpublished  
 2 (bases 1 to 2563)  
 Isogai, T. and Yamamoto, J.  
 Direct Submission  
 Submitted (20-MAR-2002) Takao Isogai, Helix Research Institute,  
 Genomics Laboratory; 1532-3 Yana, Kisarazu, Chiba 252-0812, Japan  
 (E-mail: genomics@hri.co.jp, Tel:81-438-52-3975, Fax:81-438-52-3986)  
 HRI human cDNA sequencing project; cDNA 5' & 3'-end one pass  
 sequencing, clone selection and full insert sequencing; Helix  
 Research Institute (supported by Japan Key Technology Center etc.);  
 cDNA library construction; Institute of Medical Science, University  
 of Tokyo, Laboratory of Genome Structure, Human Genome Center.  
 Location/Qualifiers  
 1. 2563  
 /organism="Homo sapiens"  
 /mol\_type="mRNA"  
 /db\_xref="taxon:9606"  
 /clone="NT2RP2005671"  
 /cell\_line="NT2"  
 /cell\_type="teratocarcinoma"  
 /clone\_lib="NT2RP2"  
 /note="cloning vector: pME18SPL3-mRNA from NT2 neuronal  
 precursor cells after 2-weeks retinoic acid (RA)  
 induction"  
 31..2241  
 /note="unnamed protein product"  
 /codon\_start=1  
 /protein\_id="BAC11608.1"  
 /db\_xref="GI:22761494"  
 /translation="MAVRALKLTLLAVVAASQAEVSEAGWGMVTPDLLFAEGTA  
 AVARGDPGVVLSMRALRRALRLRCRTQCAADFPWELDPDPSPAAGAA  
 ALRDLSPFGGLRRACLRLCPGPAHSLSEMELEFRKSPYNYLOVAYKINKLE  
 KAVAAATFPVGNPERNEMQNDYQYQMSGVKEADFKOLETPHNPFRGLRYLSE  
 EQPOEAVPHLEALQYFYFAVSEALCEGFDYDGYNLYEYNLFOAITDHYIOVL  
 NCKQNCVTELASHPSREKPPDFLPSHYNLYQPAYNYIGNYTOAVECAKTYLLFPFD

EVNQNLAYAAMLGEBHTRISIGPRESAKEVQRORSLLEKELLFFAYDVFGIPFVDPDPS  
 WTPSEVTPARQEKQKSERETAVRISQIGNLMEIETLVEBKESLDVRLRREGG  
 PLYREGISLTWNKLLNGSRQVMDGVISDECEQLRLTNVAATSGDYGQTSPT  
 PNEKFGVTVFKALKLGQSHVPLDQVLAHLYNTVEKVRIMESYFRDTPLYFGYSHL  
 VORTAIEVQAEKDDSDHPDNCILNAETLVCKVEPPAVTFRDYSAIYLNGFDG  
 GNFYTELDATKVAEVOPOCGRAVGFSSGTENPHGVKAVTRGORCAIALAFTLDPRH  
 SERDRVQADDLVKMLFSPEMDLSQEQPLDAQQGPPEPAQESLSGSSESKPKDEL"  
 ORIGIN  
 Alignment Scores:  
 Pred. No.: 0 Length: 2563  
 Score: 433.00 Matches: 733  
 Percent Similarity: 99.19% Conservative: 0  
 Best Local Similarity: 99.19% Mismatches: 3  
 Query Match: 58.83% Indels: 6  
 DB: 9 Gaps: 0  
 US-10-045-815-4 (1-736) x AK075418 (1-2563)  
 QY 1 MetAlaValAlaGAlaLeuLysLeuThrThrLeuLeuAlaValAlaAlaAlaSer 20  
 DB 31 ATGGCGGTACGCGCGTGAAGCTGCTGACACACTGCTGGCTGTGTCGCGCTGCCCTCC 90  
 QY 21 GlnAlaGluValGluSerGlnAlaGlyTrpGlyMetValThrProAspLeuLeuPheAla 40  
 DB 91 CAAGCCGAGGTCCAGTCCGAGGAGGATGGGCGATGGTACGCTGATCTGCTCTTCCGC 150  
 QY 41 GluGlyThrAlaAlaValAlaArgGlyAspTrpProGlyValValLeuSerMetGluArg 60  
 DB 151 GAGGGGACCGGACGCTACGCGCGGGGACTGGCCGGGGTGGTCTGAGCATGGAACGG 210  
 QY 61 AlaLeuArgSerArgAlaAlaLeuArgAlaLeuArgLeuArgCysArgThrGlnCysAla 80  
 DB 211 GGCTGGCTCCCGGGGAGCCCTCCGCGCCCTTGGCTGGCTGGCCACCCAGTGTGCC 270  
 QY 81 AlaAspPheProTrpGluLeuAspProAspTrpSerProSerProAlaGlnAlaSerGly 100  
 DB 271 GCCGACTTCCGCTGGGAGCTGACCCGACTGGTCCCCCAGCCGCGCCAGCCCTCGGC 330  
 QY 101 AlaGly-AlaLeuArgAspLeuSerPhePheGlyGlyLeuLeuArgAlaAlaCysLe 120  
 DB 331 GCCGC-CGCGCTGCGGACCTGAGCTTCTCGGGGGCCCTTGGCTGGCTGGCTGGCT 389  
 QY 120 uArgArgCysLeuGlyProProAlaAlaHisSerLeuSerGluMetGluLeuGluPh 140  
 DB 390 GCGCGCTGCCTCGGGCGCGCGCCGCTCGCTCAGCGAAGAGATGGAGCTGGAGTT 449  
 QY 140 eArgLysArgSerProTrpAsnTrpLeuGlnValAlaValThrPheLysIleAsnLysLeuG 160  
 DB 450 CGCAAGCGGAGCCCTACAACTACCTGCGAGTGGCTTCTCAAGATCAACAGTTGGA 509  
 QY 160 uLysAlaValAlaAlaAlaHisThrPhePheValGlyLysProGluHisMetGluMetG 180  
 DB 510 GAAAGCTGTGCTGCAGCACACACTTCTTCTGGGCAATCTCTGAGCACATGGAATGCA 569  
 QY 180 nGlnAsnLeuAspTrpTrpGlnThrMetSerGlyValLysGluAlaAspPheLysAspLe 200  
 DB 570 GCAGAACCTAGACATATTACCAAAACCATGCTCGGAGTGAAGGAGGCGCCACTTCAAGGATCT 629  
 QY 200 uGluThrGlnProHisMetGlnGluPheArgLeuGlyValArgLeuTrpSerGluGluG 220  
 DB 630 TGAGACTCAACCCCATATGCAAGAAATTCGACTGGGAGTGGAGTCTTACTCAGAGGAACA 689  
 QY 220 nProGlnGluAlaValProHisLeuGluAlaAlaLeuGlnGluTrpPheValAlaTrpG 240  
 DB 690 GCCACAGAAAGCTGTGCCCCACTAGAGCGCGCGCTGCAAGAATACTTTGTGGGCTATGA 749  
 QY 240 uGluCysArgAlaLeuCysGluGlyProTrpAspTrpAspGlyTrpAsnTrpLeuGluTy 260  
 DB 750 GGAGTGGCTGCCCTCTCGGAGGCGCCCTATGACTACGATGGCTACACTACCTTGTAGTA 809  
 QY 260 rAsnAlaAspLeuPheGlnAlaIleThrAspHisTrpIleGlnValLeuAsnCysLysG 280

810 CAACGCTGACCTCTCCAGGCCATCAGAGATCATTCACAGGTCTCAACTGTAAACA 869  
280 nAnCysValThrGluLeuAlaSerHisProSerArgGluLysProPheGluAspPheLe 300  
870 GAACTGTGTCCAGGCTTCTCTCCACCAAGTCGAGAGAGCCCTTTGAAGACTTCTT 329  
300 uProSerHisTyrAsnTyrLeuGlnPheAlaTyrTyrAsnIleGlyAsnTyrThrGlnAl 320  
930 CCATCGCATATATATCTGAGTTTGCCTACTATACATTTGGGAATATATACACAGGC 989  
320 aGly-GluCysAlaLysThrTyrLeuLeuPhePheProLeuAspGluValMetAsnGlnA 340  
990 TGT-TGAATGTGCCAGACCTATCTCTCTCTTCCCAATGACGAGGTGATGAACCAA 1048  
340 snLeuAlaTyrTyrAlaAlaMetLeuGlyGluHisThrArgSerIleGlyProArgG 360  
1049 ATTGGCTATATATGACGTATCTCTTGGAGAGAACACACAGATCATTCGGCCCCCGIG 1108  
360 luSerAlaLysGluTyrArgGlnArgSerLeuLeuGluLysGluLeuPhePheAla 380  
1109 AGAGTGCACAGGAGTACCGACAGCGAAGCCTACTGGAAAAGAACTGCTTTTCTCGCTT 1168  
380 YrAspValPheGlyIleProPheValAspProAspSerThrProGluGluValIleP 400  
1169 ATGATGTTTTTGGAAATTCCTTTTGGATCCGATTCAGGACTCCAGAGAAAGTGAATC 1228  
400 roLysArgLeuGlnGluLysGlnLysSerGluArgGluThrAlaValArgIleSerGlnG 420  
1229 CCAAGAGATTGCAAGAGAACAGAGATCAGAACGGGAAACAGCCGTACGCATCTCCAGG 1288  
420 luIleGlyAsnLeuMetLysGluIleGluThrLeuValGluGluLysThrLysGluSerL 440  
1289 AGATTGGGAACCTTATGAAGAAATCGAGACCCCTTGTGGAAGAGAAAGCAAGAGTCA 1348  
440 euAspValSerArgLeuThrArgGluGlyGlyProLeuLeuTyrGluGlyIleSerLeuT 460  
1349 TGAATGTGACACACTGACCCCGGAGGTGGCCCCCTGCTGTATGAGGCATCAGTCTCA 1408  
460 hrMetAsnSerLysLeuLeuAsnGlyTyr-GlnArgValValMetAspGlyValIleSer 479  
1409 CCATGAACCTCAAACTCCTGAATGGTTC-CCAGCGGTGTGTGATGACGGCGTAATCTCT 1467  
480 AspHisGluCysGlnGluLeuGlnArgLeuThrAsnValAlaAlaThrSerGlyAspGly 499  
1468 GACCACAGGTGTGAGGAGTGCAGAGACTGACCAATGTGCGACAGCAACCTCAGAGATGC 1527  
500 TyrArgGlyGlnThrSerProHisThrProAsnGluLysPheTyrGlyValThrValPhe 519  
1528 TACGGGGTTCAGACCTCCCCACATACTCCCAATGAAAGTTCTATGGTGTCACTGTCTTC 1587  
520 LysAlaLeuLysLeuGlyGlnGluGlyLysValProLeuGlnSerAlaHisLeuTyrTyr 539  
1588 AAAAGCCCTCAAGCTGGGGCAAGAGGCAAAATTCCTCTGCGAGAGTGCCACACCTGTACTAC 1647  
540 AsnValThrGluLysValArgArgIleMetGluSerTyrPheArgLeuAspThrProLeu 559  
1648 AACGTGACGAGAGAGTGGCGGCATCATGAGTCTTCTTCCGCTCGGTGATACGCCCTTC 1707  
560 TyrPheSerTyrSerHisLeuValCysArgThrAlaIleGluGluValGlnAlaGluArg 579  
1708 TACTTTTCTTACTCTCATCTGTGTGCGGCACTGCCATCGACAGGTCCAGGCGAGAGG 1767  
580 LysAspAspSerHisProValHisValAspAsnCysIleLeuAsnAlaGluThrLeuVal 599  
1768 AAGGATGATGTATCATCCAGTCCAGTGGCAACATGCAATCTCTGAATCCGAGAGCCCTCGTG 1827  
600 CysValLysGluProProAlaTyrThrPheArgAspTyrSerAlaIleLeuTyrLeuAsn 619  
1828 TGTGTCAAGAGCCCCCAGGCTACACCTTCGGGAGCTACAGCGCCATCTTTTACCTAAAT 1887  
620 GlyAspPheAspGlyGlyAsnPheTyrPheThrGluLeuAspAlaLysThrValThrAla 639  
1888 GGGGACTTCGATGCGGAAAATTATTTTCACTGAATGGATGCCAAGACCCGTGACGGCA 1947

QY 640 GluValGlnProGlnCysGlyArgAlaValGlyPheSerSerGlyThrGluAsnProHis 659  
Db 1948 GAGGTGACGCTTCAGTGTGGAAGAGCCGTGGATTTCTTTCAGGACACTGAAAACCCAT 2007  
QY 660 GlyValLysAlaValThrArgGlyGlnArgCysAlaIleAlaLeuTrpPheThrLeuAsp 679  
Db 2008 GGAGTGAAGGTGTCCACAGGGGGCAGCGTGTGCATCGCCCTGTGGTTTACCCTGGAC 2067  
QY 680 ProArgHisSerGluArgAspArgValGlnAlaAspAspLeuValLysMetLeuPheSer 699  
Db 2068 CCTGACACAGCGAGCGGAGCAGGGTGCAGGAGATGACCTGGTGAAGATGCTTTCACG 2127  
QY 700 ProGluGluMetAspLeuSerGlnGluGlnProLeuAspAlaGlnGlnGlyProProGlu 719  
Db 2128 CCAGAGAGATGACCTCTCCAGGAGCAGCCCTGGATGCCAGAGGGCCCCCGAA 2187  
QY 720 ProAlaGlnGluSerLeuSerGlySerGluSerLysProLysAspGluLeu 736  
Db 2188 CCTGCACAGAGTCTCTCTCAGCAGTGAATCGAAGCCCAAGGATGAGCTA 2238

## RESULT 5

AX704765 2593 bp DNA linear PAT 04-APR-2003  
LOCUS Sequence 69 from Patent WO0202634.  
DEFINITION AX704765  
ACCESSION AX704765  
VERSION AX704765.1 GI:29561431

KEYWORDS Homo sapiens (human)  
SOURCE Homo sapiens  
ORGANISM Homo sapiens

## REFERENCE

AUTHORS Gururajan,R., Hafalia,A.J., Kallick,D.A., Patterson,C., Azimzai,Y.,  
Khan,F.A., Xu,Y., Yao,M.G., Yue,H., Au-Young,J., Batza,S.,  
Baughn,M.B., Borowsky,E.A., Lo,T.P., Lu,D.A., Lu,Y., Tang,Y.T.,  
Yang,J., Zingler,K.A., Delegeane,A.M., Gietzen,K., Marcus,G.A.,  
Nguyen,D.B., Pollick,J.L., Ramkumar,J., Thangavelu,K., Wallis,N.K.  
and Warren,B.A.  
TITLE Human extracellular matrix and cell adhesion polypeptides  
JOURNAL Patent: WO 0202634-A 69 10-JAN-2002;  
INCYTE GENOMICS INC. (US)

## FEATURES

## source

1..2593 Location/Qualifiers  
/organism="Homo sapiens"  
/mol\_type="unassigned DNA"  
/db\_xref="taxon:9606"  
/note="Incyte ID No: 6780147CB1"

## ORIGIN

Alignment Scores:  
Pred. No.: 0 Length: 2593  
Score: 433.00 Matches: 733  
Percent Similarity: 99.19% Conservative: 0  
Best Local Similarity: 99.19% Mismatches: 3  
Query Match: 58.83% Indels: 6  
DB: 6 Gaps: 0

US-10-045-815-4 (1-736) x AX704765 (1-2593)

QY 1 MetAlaValArgAlaLeuLysLeuLeuThrThrLeuLeuAlaValValAlaAlaSer 20  
Db 42 ATGGCGGTACGGCGCTTGAAGCTGTGACCACTGCTGGCTGGCTGGCTGGCTGGCTGCC 101  
QY 21 GlnAlaGluValGluSerGluAlaGlyTrpGlyMetValThrProAspLeuLeuPheAla 40  
Db 102 CAAGCCGAGGTGAGTCCGAGCAGGATGGGCGATGGTGCACCTGATCTGCTCTTCGCC 161  
QY 41 GluGlyThrAlaAlaTyrAlaArgGlyAspTrpProGlyValValLeuSerMetGluArg 60  
Db 162 GAGGGGACCGGAGCTCAGCGCGGGGACCTGCGCCCGGGGTGGTCTCTGAGCATGGAACGG 221  
QY 61 AlaLeuArgSerArgAlaAlaLeuArgAlaLeuArgLeuArgCysArgThrGlnCysAla 80

Db 222 CGCTGGCTCCGGGAGCCCTCGGGCCCTCGCTGGCTGGCCGACCCAGTGTGCC 281  
Qy 81 AlaSpPheProTrpGluLeuAspProAspTrpSerProSerProAlaGlnAlaSerGly 100  
Db 282 GCCGACTTCCGCTGGAGCTGGACCCGAGCTGGTCCCGCCAGCCCGGCCAGGCTCGGGC 341  
Qy 101 AlaGly-AlaLeuArgAspLeuSerPhePheGlyGlyLeuLeuArgAlaAlaCyste 120  
Db 342 GCCGC-CGCCCTGGCGGACCTTGAGCTTCTTCGGGGGCTTCGGCTCGCCCTGCTGCT 400  
Qy 120 uArgCysLeuGlyProProAlaAlaHisSerLeuSerGluMetGluLeuGluPh 140  
Db 401 CGCGCGCTGCCGCGCGCCGCGCCCACTCGCTCAGCGAGAGATGAGCTGGAGTT 460  
Qy 140 eArgLysArgSerProTyrAsnTyrLeuGlnValAlaTyrPheLysIleAsnLysLeuG 160  
Db 461 CCGAAGCGGAGCCCTACAACTACCTGCAGGTGCGCTTCTTCAAGATCAACAAGTTGGA 520  
Qy 160 uLysAlaValAlaAlaHisThrPhePheValGlyAsnProGluHisMetGluMetG 180  
Db 521 GAAGCTGTCTGCAGCACACACTTCTTCTGGGGCAATCCTGAGCACATGGAAATGCA 580  
Qy 180 nGlnAsnLeuAspTyrTyrGlnThrMetSerGlyValLysGluAlaAspPheLysAspLe 200  
Db 581 CGAAGAACCTAGACTATTACCAACCATGTCTGGAGTGAAGGAGCGCACTTCAAGGATCT 640  
Qy 200 uGluThrGlnProHisMetGlnGluPheArgLeuGlyValArgLeuTyrSerGluGluG 220  
Db 641 TGAGACTCAACCCCATATGCAAGAAATTCGACTGGAGTGGAGTCTTCTCAGAGGAC 700  
Qy 220 rProGlnGluAlaValProHisLeuGluAlaAlaLeuGlnGluTyrPheValAlaTyrG 240  
Db 701 GCCACAGGAAGCTGTGCCCACTAGAGGCGCGCTGCAAGATATCTTTGTGGCCTATGA 760  
Qy 240 uGluCysArgAlaLeuCysGluGlyProTyrAspTyrAspGlyTyrAsnTyrLeuGluTy 260  
Db 761 GGAGTGGCGTCCCTCTCGAAGGGCCCTATGACTAGATGGCTTACCACTTCTTGAGTA 820  
Qy 260 rAsnAlaAspLeuPheGlnAlaIleThrAspHisTyrIleGlnValLeuAsnCysLysG 280  
Db 821 CAACGCTGACCTCTCCAGGCCATCACAGATCATATCATCCAGTCTCCAACTGTAAACA 880  
Qy 280 rAsnCysValThrGluLeuAlaSerHisProSerArgGluLysProPheGluAspPheLe 300  
Db 881 GAACCTGTGTCCGAGCTTGTCTCCCAACCAAGTCGAGAGAGCCCTTTGAAGACTTCCT 940  
Qy 300 uProSerHisTyrAsnTyrLeuGlnPheAlaTyrTyrAsnIleGlyAsnTyrThrGlnAl 320  
Db 941 CCCATCGCATTAATATATCTGCGAGTTGCTACTATTAATTTGGGAAATATACAGGC 1000  
Qy 320 aGly-GluCysAlaLysThrTyrLeuLeuPhePheProAsnAspGluValMetAsnGlnA 340  
Db 1001 TGT-TGAATGTGCCAAGACCTATCTTCTTCTTCCCAATGACGAGGTGATGAACCAAA 1059  
Qy 340 snLeuAlaTyrTyrAlaAlaMetLeuGlyGluGluHisThrArgSerIleGlyProArg 360  
Db 1060 ATTTGGCCCTATATGCACTATGCTTGAGAGAACACACAGATCATTCGCGCCCGCTG 1119  
Qy 360 luSerAlaLysGluTyrArgGlnArgSerLeuLeuGluLysGluLeuLeuPhePheAla 380  
Db 1120 AGAGTGCCAGAGGTACCGACAGCGAGCCCTTCTGGAAAAGAACTGCTTTTCTGCTT 1179  
Qy 380 yrAspValPheGlyIleProPheValAspProAspSerTrpThrProGluGluValIleP 400  
Db 1180 ATGATGTTTTGGAAATTTCCCTTTTGGATTCGGATTATGACTTCCAGAAAGATGATTC 1239  
Qy 400 roLysArgLeuGlnGluLysGlnLysSerGluArgGluThrAlaValArgIleSerGlnG 420  
Db 1240 CCAAGAGATTCCAAGAGAAACAGAAAGTCAGAACCGGAAACAGCCGTACGCATCTCCAGG 1299  
Qy 420 luIleGlyAsnLeuMetLysGluIleGluThrLeuValGluGluLysThrLysGluSerL 440

Db 1300 AGATTGGGAACCTTATGAAGAAATCGAGACCCCTTGTGGAAGAGAACCAAGAGTCA 1359  
Qy 440 euAspValSerArgLeuThrArgGluGlyGlyProLeuLeuTyrGluGlyIleSerLeuT 460  
Db 1360 TGGATGTGAGCAGACTGACCCGGGAAGTGGCCCTCTGTATGAAGGCATAGTCTCA 1419  
Qy 460 hnMetAsnSerLysLeuLeuLeuAsnGlyTyr-GlnArgValValMetAspGlyValIleSer 479  
Db 1420 CCATGAACTCCAACTCTCTGAATGTTCC-CAGCGGGTGGTATGGAGCGGGTAACTCT 1478  
Qy 480 AspHisGluCysGlnGluLeuGlnArgLeuThrAsnValAlaAlaThrSerGlyAspGly 499  
Db 1479 GACCAAGAGTGTGAGAGCTGCAGAGACTGACCAATGTGGCAGCAACCTCAGGAGATGC 1538  
Qy 500 TyrArgGlyGlnThrSerProHisThrProAsnGluLysPheTyrGlyValThrValPhe 519  
Db 1539 TACCGGGTGCAGACTCCCCACATCTCCCAATGAAGAAGTCTATGGTGTCACTGTCTTC 1598  
Qy 520 LysAlaLeuLysLeuGlyGlnGluLysValProLeuGlnSerAlaHisLeuTyrTyr 539  
Db 1599 AAAGCCCTCAAGCTGGGCAAGAGCAAGTTCTCTGCAGAGTGCCTGTACTAC 1658  
Qy 540 AsnValThrGluLysValArgArgIleMetGluSerTyrPheArgLeuAspThrProLeu 559  
Db 1659 AACGTGACGAGAGAGTGGCGCATCATGAGTCTTCTTCCGCTGGATACGCCCTC 1718  
Qy 560 TyrPheSerTyrSerHisLeuValCysArgThrAlaIleGluGluValGlnAlaGluArg 579  
Db 1719 TACTTTCTCTCTCATCTGCTGTGCGCACTGCATCGAAGAGTCCAGCAGAGAGG 1778  
Qy 580 LysAspAspSerHisProValHisValAspAsnCysIleLeuAsnAlaGluThrLeuVal 599  
Db 1779 AAGGATGATAGTATCTCCAGTCCAGTGCAGCAACTGCATCCTGAATGCCGAGACCTGTG 1838  
Qy 600 CysValLysGluProProAlaTyrThrPheArgAspTyrSerAlaIleLeuTyrLeuAsn 619  
Db 1839 TGTGTCAAGAGCCCGCAGCTTACACTTCGCGACTACAGCGCATCTTTTACCTAAAT 1898  
Qy 620 GlyAspPheAspGlyGlyAsnPheTyrPheThrGluLeuAspAlaLysThrValThrAla 639  
Db 1899 GGGGACTTCGATGGCGAAACTTTTATTTCTCACTGAATCGATGCCAAGCCGTGACGCA 1958  
Qy 640 GluValGlnProGlnCysGlyArgAlaValGlyPheSerSerGlyThrGluAsnProHis 659  
Db 1959 GAGGTGACGCTCAGTGTGGAAGAGCCGTGGATTTCTTTCAGGCATCGAAACCCACAT 2018  
Qy 660 GlyValLysAlaValThrArgGlyGlnArgCysAlaIleAlaLeuTyrPheThrLeuAsp 679  
Db 2019 GGAGTGAAGCTGTCAACAGGGGCGAGCTGTGCTATCGCCCTGTGGTTTCACTGGAC 2078  
Qy 680 ProArgHisSerGluArgAspArgValGlnAlaAspLeuValLysMetLeuPheSer 699  
Db 2079 CTTGACACACAGCGAGCGGAGTGGAGTGCAGGAGTACCTGGTGAAGATGCTTCTCAGC 2138  
Qy 700 ProGluGluMetAspLeuSerGlnGluProLeuAspAlaGlnGlnGlyProProGlu 719  
Db 2139 CAGAGAGATGAGCTCTCCAGAGGAGCCCTGGATGCCAGAGGGCCCCCGAA 2198  
Qy 720 ProAlaGlnGluSerLeuSerGlySerGlySerLysProLysAspGluLeu 736  
Db 2199 CCTGCACAGAGTCTCTCTCGGCAAGTGAATCGAAGCCCAAGAGTGAAGCTA 2249

## RESULT 6

AX877288  
LOCUS AX877288 2595 bp DNA linear PAT 17-DEC-2003  
DEFINITION Sequence 12193 from Patent EP1074617.  
ACCESSION AX877288  
VERSION AX877288.1 GI:40032024  
KEYWORDS Homo sapiens (human)  
SOURCE Homo sapiens  
ORGANISM Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE 1  
 AUTHORS Ota, T., Isozaki, T., Nishikawa, T., Hayashi, K., Saito, K., Yamamoto, J., Ishii, S., Sugiyama, T., Wakamatsu, A., Nagai, K. and Otsuki, T.  
 TITLE Primers for synthesizing full-length cDNA and their use  
 JOURNAL Patent: EP 1074617-A 12193 07-FEB-2001;  
 Research Association for Biotechnology (JP)  
 FEATURES  
 source 1. .2585  
 /organism="Homo sapiens"  
 /mol\_type="unassigned DNA"  
 /db\_xref="taxon:9606"

## FEATURES

1. .2585  
 /organism="Homo sapiens"  
 /mol\_type="unassigned DNA"  
 /db\_xref="taxon:9606"

## CDS

53. .2263  
 /notes="unnamed protein product"  
 /codon\_start=1  
 /protein\_id="CAE89658.1"  
 /db\_xref="GI:40032025"  
 /translation="MAVRALKLLTLLVAARASQAEVSEAGMGVTPDLLPAECTA  
 AYARGWPQVILSMERALSRAALRALRLRCRTQCAADFPWELDPWSFSPAQASAA  
 ALKDLFFGLLRRAELRCGLDYPAAHSGSEMELEFRKRSPYNTLQVAYFKINKLE  
 KAVAAHFFEVGPEMEVQNDLYYTWSGKEADFQKOLETPHMQEFLRVLYSE  
 EQPQEAHPLEALQEFYVAYEACRALCEGPDYDGYNYLEVNADLFQAITDHYIOVL  
 NCKONCVTEALSHPSREKPEDEFLPSHYNYLQPAYVNYGNVQAVECAKTYLLFFPD  
 EVNQNOLYVAAMLEHRTS IGPRESAKEYRORSLEKELLFFAYDVFGI PFVDPDS  
 WTEPEVLPKLEQKQSERETA VLSQELGNLMKEIETLVESKTESLDVDSLTREGG  
 PLTEGISLTMSKLNLSORVNDGVISDHCEQELQRLTNVATSGDGYRQTSPT  
 PNEKFGVTVFKALKLQEGKVPQLQSAHLINVTETKVRIMESYFRLDIPLFYSSHL  
 VCFATLEEVQAEKDDSHFVNDNCINLAETLVCKEPPAYTFRDSALYLNQDFDG  
 GNFYFTELDKTVTAEVQPOCGRAVGFSSGTENPHGVKAVTRQRCALIAWFTLDPRH  
 SERDRVQADLVKMLFSPEDMLSQEQLDAQQGPPEPAQESLSGESKPKDEL"

## ORIGIN

Alignment Scores:  
 Pred. No.: 0 Length: 2585  
 Score: 433.00 Matches: 733  
 Percent Similarity: 99.19% Conservative: 0  
 Best Local Similarity: 99.19% Mismatches: 3  
 Query Match: 58.83% Indels: 6  
 DB: 6 Gaps: 0

US-10-045-815-4 (1-736) x AX877288 (1-2585)

QY 1 MetAlaValArgAlaLeuLysLeuLeuThrThrLeuLeuAlaValValAlaAlaLaser 20  
 Db 53 ATGGCGGTACGGCGGTGAGCTGTGACACACACTGCTGCTGCTGCTGCTGCTGCTGCTC 112  
 QY 21 GlnAlaGluValGluSerGluAlaGlyTyrGlyMetValThrProAspLeuLeuPheAla 40  
 Db 113 CAAGCCGAGGTGAGTCCGAGGAGGATGGGCAATGGTGCCTGATCTGCTCTTGGCC 172  
 QY 41 GluGlyThrAlaAlaTyrAlaAvtGlyAspTyrProGlyValValLeuSerMetGluArg 60  
 Db 173 GAGGGGACCGAGCTACGCGCGGGGACTGGCCCGGGGTGGTCTCTGAGCATGGAAACGG 232  
 QY 61 AlaLeuArgSerArgAlaAlaLeuArgAlaLeuArgLeuArgCysArgThrGlnCysAla 80  
 Db 233 CGCGTGGCTCCCGGGCAGCCCTCCGCGCTTCCGCTGCTGCTGCTGCTGCTGCTGCTGCTC 292  
 QY 81 AlaAspPheProTyrGluLeuAspProAspTyrSerProAlaGlnAlaSerGly 100  
 Db 293 GCCCACTTCCCGTGGGAGCTGAGACCCGACTGTGTCTCCCGCCGCGCCGAGGCTCGGGC 352  
 QY 101 AlaGly-AlaLeuArgAspLeuSerPhePheGlyLeuLeuArgArgAlaAlaCysLe 120  
 Db 353 GCCGC-CGCCCTGGCGACCTGAGCTTCTTCGGGGGCTTCTGCTGCTGCTGCTGCTGCT 411  
 QY 120 uArgArgCysLeuGlyProProAlaAlaHisSerLeuSerGluGluMetGluLeuGluPh 140  
 Db 412 CGCGCGGTGCTCGGGCGCGCGCGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 471  
 QY 140 eArgLysArgSerProTyrAsnTyrLeuGlnValAlaTyrPheLysIleAsnLysLeuGlu 160  
 Db 472 CCGCAAGCGGAGCCCTTACAACTACCTGAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 531

QY 160 uLysAlaValAlaAlaAlaHisThrPhePheValGlyAsnProGluHisMetGluMetGlu 180  
 Db 532 GAAAGCTGTGCTGCGAGCACACACCTTCTTCTGCGGCAATCTCTGAGCACATGGAATGCA 591  
 QY 180 nGlnAsnLeuAspTyrTyrGlnThrMetSerGlyValLysGluAlaAspPheLysAspLe 200  
 Db 592 GCAGAACCTAGACTATTACCAAAACATGTCTGGAGTGAAGGAGGCGGCTTCAAGAGTCT 651  
 QY 200 uGluThrGlnProHisMetGlnGluPheArgLeuGlyValArgLeuTyrSerGluGluGlu 220  
 Db 652 TGAGACTCAACCCCATATGCAAGAAATTTGAGTGGAGTGGGACTCTACTCAGAGGAACA 711  
 QY 220 nProGlnGluAlaValProHisLeuGluAlaAlaLeuGlnGluTyrPheValAlaTyrGlu 240  
 Db 712 GCCACAGGAAGCTGTGCCCCACCTAGAGGCGGCTGCAAGATACTTTGTGGCCTATGA 771  
 QY 240 uGluCysArgAlaLeuCysGluGlyProTyrAspTyrAspGlyTyrAsnTyrLeuGluTyr 260  
 Db 772 GGAGTGGCTGCTTCTGCGAAGGCCCTATGACTACGATGGCTACACTACCTTGAGTA 831  
 QY 260 rAsnAlaAspLeuPheGlnAlaIleThrAspHisTyrIleGlnValLeuAsnCysLysGlu 280  
 Db 832 CAACGCTGACCTCTTCCAGGCCATCACAGATCATTACATCCAGGTCTCTCAACTGTGAACA 891  
 QY 280 rAsnCysValThrGluLeuAlaSerHisProSerArgGluLysProPheGluAspPheLe 300  
 Db 892 GAACGTGTGTCGAGAGCTGTCTTCCCAAGTCGAGAGAGCCCTTTGAAGACTTCTCT 951  
 QY 300 uProSerHisTyrAsnTyrLeuGlnPheAlaTyrTyrAsnIleGlyAsnTyrThrGlnAla 320  
 Db 952 CCCATCGCATTAATATCTGCGCTTCTGCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCT 1011  
 QY 320 aGly-GluCysAlaLysThrTyrLeuLeuPhePheProAsnAspGluValMetAsnGlnAla 340  
 Db 1012 TGT-TGAATGTGCCAAGACCTATCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCT 1070  
 QY 340 snLeuAlaTyrTyrAlaAlaMetLeuGlyGluGluHisThrArgSerIleGlyProAspG 360  
 Db 1071 ATTGGCTTATATGACGCTATGCTTGGAGAAGAACACACAGATCCATCGGCCCCCGTG 1130  
 QY 360 luSerAlaLysGluTyrArgGlnArgSerLeuLeuGluLysGluLeuLeuPhePheAlaTyr 380  
 Db 1131 AGAGTGCCAAAGAGTACCGACAGCGAAGCTACTGGAAGAAAGAACTGCTTCTTCTTCTTCTTCT 1190  
 QY 380 rAspValPheGlyIleProPheValAspProAspSerTyrThrProGluGluValIleP 400  
 Db 1191 ATGATGTTTTTGGAAATTCCTTTTGGATTCGGATTCGGATTCGGATTCGGATTCGGATTC 1250  
 QY 400 rLysArgLeuGlnGluLysGlnLysSerGluArgGluThrAlaValArgIleSerGlnG 420  
 Db 1251 CCAAGAGATTGCAAGAGAAACAGAACTCAGAAACGGGAAACAGCCGTACGCATCTCCACGG 1310  
 QY 420 luIleGlyAsnLeuMetLysGluIleGluThrLeuValGluGluLysThrLysGluSerL 440  
 Db 1311 AGATTTGGGAACCTTATGAAGAAATCGAGACCTTGTGGAAGAGAACCAAGGAGTCTAC 1370  
 QY 440 euAspValSerArgLeuThrArgGluGlyGlyProLeuLeuTyrGluGlyIleSerLeuT 460  
 Db 1371 TGGATGTGACAGACTGACCCGGGAGGTGGGCCCTTCTGCTGATGAGGCATCAGTCTCA 1430  
 QY 460 hrMetAsnSerLysLeuLeuGlnGlyTyr-GlnArgValValMetAspGlyValIleSer 479  
 Db 1431 CCATGAACCTCCAAACTCTCTGAATGGTTC-CCAGCGGTGGTGGATGGACGGCTAATCTCT 1489  
 QY 480 AspHisGluCysGlnGluLeuGlnArgLeuThrAsnValAlaAlaThrSerGlyAspGly 499  
 Db 1490 GACACAGAGTGTGAGAGCTGCAGAGCTGACCAATGTGGCAGCAACCTCAGGAGATGGC 1549  
 QY 500 TyrArgGlyGlnThrSerProHisThrProAsnGluLysPheTyrGlyValThrValPhe 519  
 Db 1550 TACCGGGGTGACAGCTCCCACTATCTCCCAATGAAAGTCTCTATGGTGTCACTGTCTTC 1609  
 QY 520 LysAlaLeuLysLeuGlyGlnGluGlyLysValProLeuGlnSerAlaHisLeuTyrTyr 539





832 CAACGCTGACCTCTCCAGCGCCATCACAGATCATTTACATCCAGGTCTCTCAACTGTAAACA 891  
Qy 280 nAnCysValThrGluLeuAlaSerHisProSerArgGluLeuPheProPheLeuAspPheLe 300  
Db 892 GAATGTGTCCAGGAGCTTCTCCACCCCAAGTCGAGAGAACCCCTTTGAAGACTTCCT 951  
Qy 300 uProSerHisTyrAsnTyrLeuGlnPheAlaTyrTyrAsnIleGlyAsnTyrThrGlnAl 320  
Db 952 CCCATCGCATTAATAATTATCTGCAGTTTGCCTACTATAACATTGGGAATTATACACAGGC 1011  
Qy 320 aGly-GluCysAlaLysThrTyrLeuLeuPhePheProAsnAspGluValMetAsnGlnA 340  
Db 1012 TGT-TGAATGTGCCAAGACCTATCTCTCTCTTCCCAATGACGAGGTGATGAACAA 1070  
Qy 340 snLeuAlaTyrTyrAlaAlaMetLeuGlyGluGlnHisThrArgSerIleGlyProArg 360  
Db 1071 ATTTGGCTTATTATGACGCTATGCTTGGAGAAGACACACCATCCATCCATCGGCCCGGTG 1130  
Qy 360 luSerAlaLysGluTyrArgGlnArgSerLeuLeuGlnLysGluLeuLeuPhePheAla 380  
Db 1131 AGAGTGCACAGGAGTACCGACAGCGAAGCCCTACTGGAAAAAGAACTGCTTTCTTCGCTT 1190  
Qy 380 YrAspValPheGlyIleProPheValAspProAspSerTyrThrProGluGluValIleP 400  
Db 1191 ATGATGTTTTTGGAAATTCCTTTGTGGATCCGATTCATGCAATCCAGAAAGTGAATC 1250  
Qy 400 roLysArgLeuGlnGluLysGlnLysSerGluArgGluThrAlaValArgIleSerGlnG 420  
Db 1251 CCAAGAGATTGCAAGAGAAACAGAGAGTCAGAACGGGAAACAGCGGTACGCATCTCCACG 1310  
Qy 420 lulleGlyAsnLeuMetLysGluIleGluThrLeuValGluGluLysThrLysGluSerL 440  
Db 1311 AGATTGGGAACCTTATGAAGAAATCGAGACCCCTTGTGGAGAGAGAACCAAGAGGTAC 1370  
Qy 440 euAspValSerArgLeuThrArgGluGlyProLeuLeuTyrGluGlyIleSerLeu 460  
Db 1371 TGGATGTGACGAGACTGACCGGAGAGTGGCCCTGCTGTATGAAGGATCAGTCTCA 1430  
Qy 460 hrMetAsnSerLysLeuLeuAsnGlyTyr-GlnArgValValMetAspGlyValIleSer 479  
Db 1431 CCATGAACCTCAAACTCCTGAATGGTTC-CCAGCGGGTGTGTGATGACGCGCTAATCTCT 1489  
Qy 480 AspHisGluCysGlnGluLeuGlnArgLeuThrAsnValAlaAlaThrSerGlyAspGly 499  
Db 1490 GACCACAGGTGTACGAGAGTGCAGAGACTGACCAATGTGCAGCAACCTCAGAGAGTGC 1549  
Qy 500 TyrArgGlyGlnThrSerProHisThrProAsnGluLysPheTyrGlyValThrValPhe 519  
Db 1550 TACCGGGTTCAGACCTCCCCACATACCTCCCAATGAAAGTTCTATGGTGTCACTGTCTC 1609  
Qy 520 LysAlaLeuLysLeuGlyGlnGluLysValProLeuGlnSerAlaHisLeuTyrTyr 539  
Db 1610 AAAGCCCTCAAGCTGGGGCAAGAGGCAAGTTCTCTGAGAGTGGCCACCTGTACTAC 1669  
Qy 540 AsnValThrGluLysValArgAtGileMetGluSerTyrPheArgLeuAspThrProLeu 559  
Db 1670 AACGTGACGAGAGAGTGGCGGCGATCATGAGTCTTACTTCGCTGGATCGCCCTC 1729  
Qy 560 TyrPheSerTyrSerHisLeuValCysArgThrAlaIleGluGluValGlnAlaGluArg 579  
Db 1730 TACTTTTCTACTCTCATCTGTGTGCGCACTGCCATCGAAGAGGTCCAGGCGAGAGG 1789  
Qy 580 LysAspAspSerHisProValHisValAspAsnCysIleLeuAsnAlaGluThrLeuVal 599  
Db 1790 AAGGATGATGATCATCCAGTCCAGTCCAGTCCAGTCCATCTGATGCGGAGACCTCGTG 1849  
Qy 600 CysValLysGluProProAlaTyrThrPheArgAspTyrSerAlaIleLeuTyrLeuAsn 619  
Db 1850 TGTGTCAAAGAGGCCCCAGCGCTACACCTTCGCGACTACAGCGGCATCTTTACCTAAAT 1909  
Qy 620 GlyAspPheAspGlyGlyAsnPheTyrPheThrGluLeuAspAlaLysThrValThrAla 639  
Db 1910 GGGGACTTCGATGCGGAAACTTTTATTCTACTGAATGGATGCAAGACCGTGACGGCA 1969

Qy 640 GluValGlnProGlnCysGlyArgAlaValGlyPheSerSerGlyThrGluAsnProHis 659  
Db 1970 GAGGTGAGCCCTCAGTGTGGAGAGCCCTGGGATCTCTTTCAGGACTGAAAACCCCAT 2029  
Qy 660 GlyValLysAlaValThrArgGlyGlnArgCysAlaIleAlaLeuTyrPheThrLeuAsp 679  
Db 2030 GGAGTGAAGGCTGTCCACAGGGGGCAGGCTGTGCATCGCCCTGTGGTTTACCCTGGAC 2089  
Qy 680 ProArgHisSerGluArgAspArgValGlnAlaAspAspLeuValLysValLeuPheSer 699  
Db 2090 CTTCCACACAGCGGACGGGATGAGGTGAGGAGATGACCTGTGTGAAGATGCTCTTCAGC 2149  
Qy 700 ProGluGluMetAspLeuSerGlnGlnProLeuAspAlaGlnGlnGlyProProGlu 719  
Db 2150 CCAGAAGAGATGAGCTCTCCAGGAGCAGCCCTGTGATGCCAGCAGGGCCCCCGCAA 2209  
Qy 720 ProAlaGlnGluSerLeuSerGlySerGluSerLysProLysAspGluLeu 736  
Db 2210 CCTGCACAAGAGTCTCTCTCAGGCAGTGAATCGAAGCCCAAGGATGAGCTA 2260

## RESULT 8

AK027697 2585 bp mRNA linear PRI 01-AUG-2002  
Homo sapiens cDNA FLJ14791 fis, clone NT2RP4001064, weakly similar  
to SYNAPTONEMAL COMPLEX PROTEIN SC65.

AK027697

AK027697.1 GI:14042570

oligo capping; fis (full insert sequence).

Homo sapiens (human)

Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

## REFERENCE

## AUTHORS

Isogai,T., Ota,T., Hayashi,K., Sugiyama,T., Otsuki,T., Suzuki,Y.,  
Nishikawa,T., Nagai,K., Sugano,S., Aotsuka,S., Yoshikawa,Y.,  
Matsumawa,H., Ishii,S.; Kawai,Y., Saito,K., Yamamoto,J.,  
Wakamatsu,A., Nakamura,Y., Nagahari,K., Masuho,Y. and Sasaki,N.  
NEDO human cDNA sequencing project

Unpublished

2 (bases 1 to 2585)

Isogai,T. and Otsuki,T.

Direct Submission

Submitted (10-MAY-2001) Takao Isogai, Helix Research Institute,  
Genomics Laboratory; 1532-3 Yana, Kisarazu, Chiba 292-0812, Japan  
(E-mail:genomics@hri.co.jp, Tel:81-438-52-3975, Fax:81-438-52-3986)  
NEDO human cDNA sequencing project supported by Ministry of  
Economy, Trade and Industry of Japan; cDNA full insert sequencing;  
Research Association for Biotechnology; cDNA library construction;  
5'- & 3'-end one pass sequencing and clone selection; Helix  
Research Institute (supported by Japan Key Technology Center etc.)  
and Department of Virology, Institute of Medical Science,  
University of Tokyo.

## FEATURES

## source

1..2585  
/organism="Homo sapiens"  
/mol\_type="mRNA"  
/db\_xref="taxon:9606"  
/clone="NT2RP4001064"  
/cell\_line="NT2"  
/cell\_type="teratocarcinoma"  
/clone\_lib="NT2RP4"  
/note="cloning vector: pME18SFL3-mRNA from NT2 neuronal  
precursor cells after 2-weeks retinoic acid (RA)  
induction."  
53..2263  
/note="unnamed protein product"  
/codon\_start=1  
/protein\_id="BAB55305.1"  
/db\_xref="GI:14042571"

## CDS

translation="NAVRALKLTLLAVAAASOAEVSEBAGWMVTPDLLFAEGTA  
AYARGDWQCVLSMERALRSRALPRLRCFTQCAADPHELPDQSPSPQAQSGAA  
ALRDLUSFFGLLRRAACLRRCGLGPPAHSLEMELEFRKSPNYLQVAFKINKLE



KAVAAATFFVGNPEHMEMQNLDDYQTMSCVKEADFKDLETQPHMQEFLGLVRLISE  
 EPOEAVPHLEAALQEFYVEECRALCEBGYDDGYNLEYNADLFQALDHYIQVL  
 NCKQCVLEASHSRKPFEDFLPSHYNYLOFAYNIGNYQAVECACTYLLFFPND  
 EVNQNLAYAAAMEGEHETRSIGPESAKERYQSRLEKLLFFAYDVFIPFVDPDS  
 WTPREVIPKRLQEKQKESRETAIVISOEIGNLMEKEITLVEEKSLDVSRLTRREG  
 PLLYEGISLTWNSKLNGSORVMDGVISDHECEQELORLTNVAATSDGVYRGOTSPHT  
 PNEKPYGVTFKALGOEGKVPLQSAHLVYNTKRYRIMESVRLDTPLYSYSHL  
 VCRFAIEVQAKRQDSDHPVNDVNCINLAETLVCKBPAYTFNDSAILYLNGDSDG  
 GNIFYTEDAKNTVAEVOQCGRVAFSSFTENPHGVKAVTRGORCAIALWFLDPRH  
 SERDRVQADDILVKMLFSPPEMDLSQEPDLQAQGPPEPAQESLSGSSSEKPKDEL"

## ORIGIN

## Alignment Scores:

Pred. No.: 0 Length: 2585  
 Score: 433.00 Matches: 733  
 Percent Similarity: 99.19% Conservative: 0  
 Best Local Similarity: 99.19% Mismatches: 3  
 Query Match: 58.83% Indels: 6  
 DB: 9 Gaps: 0

US-10-045-815-4 (1-736) x AK027697 (1-2585)

QY	260	rAsnAlaAspLeuPheGlnAlaIleThrAspHisTyrIleGlnValLeuAsnCysLysG	280
Db	832	CAACGCTGACCTCTTCCAGCGCATCACAGATCATTTACATCCAGGTCCTCACTGTAAGCA	891
QY	280	nAsnCysValThrGluLeuAlaSerHisProSerArgGluLysProPheGluAspPheLe	300
Db	892	GAACTGTGTACGGAGCTTGTCTCCACCAGTCCAGAGAGCCCTTTGAAGACTTCT	951
QY	300	uProSerHisTyrAsnTyrLeuGlnPheAlaTyrTyrAsnIleGlyAsnTyrThrGlnAl	320
Db	952	CCCATCGCATTAATAATTATCTGCAGTTTGCCTACTATAACATGGGAATTATACACGGC	1011
QY	320	aGly-GluCysAlaLysThrTyrLeuLeuPhePheProAsnAspGluValMetAsnGlnA	340
Db	1012	TGT-TGAATGTGCCAAGACCTATCTCTCTCTCCCAATGACGAGGTGATGAACCAA	1070
QY	340	snLeuAlaTyrTyrAlaAlaMetLeuGlyGluGluHisThrArgSerIleGlyProArgG	360
Db	1071	ATTGGCCTATTATGAGCTATGCTTGGAGAAGACACACAGATCCATCGCCCGCTG	1130
QY	360	luSerAlaLysGluTyrArgGlnArgSerLeuLeuGluLysGluLeuLeuPheAlaI	380
Db	1131	AGAGTGCACAGAGTACCGACAGCGAAGCTACTCGAAAAAGAACTGCTTTCTTCGTT	1190
QY	380	YrAspValPheGlyIleProPheValAspProAspSerTyrThrProGluGluValIleP	400
Db	1191	ATGATGTTTGGANTTCCCTTTGTTGGATCGGATTCATGGACTCCAGAGAGTGATTC	1250
QY	400	roLysArgLeuGlnGluLysGlnLysSerGluArgGluThrAlaValArgIleSerGlnG	420
Db	1251	CCAAGAGATTGCAAGAGAAACAGAAAGTCAGAACCGGAAACAGCCGTACGCATCTCC	1310
QY	420	luIleGlyAsnLeuMetLysGluIleGluThrLeuValGluLysThrLysGluSerL	440
Db	1311	AGATTGGGAACCTTATGAAGAAATCGAGACCTTGTGGAAGAGAACCAAGAGTGC	1370
QY	440	euAspValSerArgLeuThrArgGluGlyGlyProLeuLeuTyrGluGlyIleSerLeu	460
Db	1371	TGGATGTGAGCAGACTGACCCGGGAAGGTGGCCCTCTGTATGAAGGCATCAGTCTCA	1430
QY	460	hrMetAsnSerLysLeuLeuAsnGlyTyr-GlnArgValValMetAspGlyValIleSer	479
Db	1431	CATGAACCTCCAACTCCTGAATGGTTC-CCAGCGGGTGGTGGTGGAGCGCGTAACTCT	1489
QY	480	AspHisGluCysGlnGluLeuGlnArgLeuThrAsnValAlaAlaThrSerGlyAspGly	499
Db	1490	GACCAAGAGTGTACAGAGCTGACCAATGTGGCAGCAACCTCAGGAGATGGC	1549
QY	500	TyrArgGlyGlnThrSerProHisThrProAsnGluLysPheTyrGlyValThrValPhe	519
Db	1550	TACCGGGGTGAGACCTCCCCACATCTCCCAATGAAAGTTCTATGGTGTCACTGTCTTC	1609
QY	520	LysAlaLeuLysLeuGlyGlnGlyLysValProLeuGlnSerAlaHisLeuTyrTyr	539
Db	1610	AAAGCCCTCAGCTGGGCAAGAGCAAGTCTCTCTGAGAGTGGCCACCTGTACTAC	1669
QY	540	AsnValThrGluLysValArgArgIleMetGluSerTyrPheArgLeuAspThrProLeu	559
Db	1670	AAAGTGTGAGGAGAGGTGGCGCATCATGAGTCTCTTCCGCTGGATACGCCCTC	1729
QY	560	TyrPheSerTyrSerHisLeuValCysArgThrAlaIleGluValGlnAlaGluArg	579
Db	1730	TACTTTTCTACTCTCACTGTGTGGCGCACTGCCATCGAAGAGGTCCAGCAGAGAGG	1789
QY	580	LysAspAspSerHisProValHisValAspAsnCysIleLeuAsnAlaGluThrLeuVal	599
Db	1790	AAGGATGATGATCATCCAGTCCAGTGGACAACTGCATCTCTGAATGCCAGACCTCGTG	1849
QY	600	CysValLysGluProProAlaTyrThrPheArgAspTyrSerAlaIleLeuTyrLeuAsn	619
Db	1850	TGTGTCAAAGAGGCCCCAGCCTTACACCTTCGCGACTACAGCGCCATCTTACTTAAT	1909

```
QY 620 GlyAspPheAspGlyGlyAsnProPheThrGluLeuAspAlaValThrValAla 639
Db 1910 GGGGACTTCGATCGCGGAACTTTTATTTCACTGAACCTGGATGCCAAGACCGTGACGGCA 1969
QY 640 GluValGlnProGlnCysGlyArgAlaValGlyPheSerSerGlyThrGluAsnProHis 659
Db 1970 GAGGTGAGCCTCAGTGTGGAAGCGGTGGATTTCTTTCAGGCCTGAARACCCACAT 2029
QY 660 GlyValLysAlaValThrArgGlyGlnArgCysAlaValLeuThrPheThrLeuAsp 679
Db 2030 GGAGTGAAGGTGTTCACAGGGGCGACGCTGTGCCATCCCTGTGGTTTCACTGGAC 2089
QY 680 ProArgHisSerGluArgAspArgValGlnAlaAspAspLeuValLysMetLeuPheSer 699
Db 2090 CCTCGACACAGCGAGCGGACAGGGTGCAGGACGATGACCTGTGAGATGCTCTTCAGC 2149
QY 700 ProGluGlnMetAspLeuSerGlnGlnProLeuAspAlaGlnGlnGlyProProGlu 719
Db 2150 CCAGAGAGATGACCTCTCCAGGAGCAGCGCTGGATGCCAGCAGGCGCCCGCGAA 2209
QY 720 ProLagInGluSerLeuSerGlySerGluSerLysProLysAspGluLeu 736
Db 2210 CCTGCACAAGAGTCTCTCAGGCAGTGAATCGAAGCCCAAGGATGAGCTA 2260

RESULT 9
LOCUS AX877139 2524 bp DNA linear PAT 17-DEC-2003
DEFINITION Sequence 12044 from Patent EP1074617.
ACCESSION AX877139
VERSION AX877139.1 GI:40031875
KEYWORDS
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1
AUTHORS Ota,T., Isogai,T., Nishikawa,T., Hayashi,K., Saito,K., Yamamoto,J.,
Ishii,S., Sugiyama,T., Wakamatsu,A., Nagai,K. and Otsuki,T.
TITLE Primers for synthesising full-length cDNA and their use
JOURNAL Patent: EP 1074617-A 12044 07-FEB-2001;
Research Association for Biotechnology (JP)
FEATURES
source
1. 2524
/organism="Homo sapiens"
/mol_type="unassigned DNA"
/db_xref="taxon:9606"
16_2226
/feature="unassigned protein product"
/codon_start=1
/db_xref="GI:40031876"
/translation="MAVRALKLLTLLAVVAASQAEVESEAGWGVTPDLLFAEGTA
AYRGDWPVGLSMEALRSLRALRLRCRTQCAADFPWELDPDWSPPSAQSGAA
ALRDLSPFGGLREAAALRCLGPPAAHSISEMELEFRKRSFVNYLOVAFYKINLE
KAVAAATFFVGNPEHEMEXQNDLYCTNSGVKEADFKLETOPHMOEFLGVLYSE
EQQEAAPHLAALQEFVAYEECALCEGPDYDNGINLYSTNADLFQAITDHYIQL
NQNQCVTELASRPSREKPFDFLPSHYNLYQFAYINIGNYQAVECAKTLFFPND
EWMQNAYLAAMLGEEHRSIGRESAYKRSLEKELLFAYDVFGLPFPDPS
WTPBEVTPKRLQEKQKSERETAVRISQEIGNLMKEIETLVEEKTESLDVSRTRBGG
PLLYEG:SLTNWSQSRVMDGVIISDHECOELRLTNVAATSDGVRGQTSFPH
PNEKFGVTFKALKIQEGKVPQSAHLYNNTKVRIMESVFLDTPLYFSYSHL
VCRTAIEVQAEKDDSHPHVDNCLINAEILNCKEPPATPRDYSAIILANGDFDG
GNFRTFELDAKTVAIEVQPCGRAVFGSSGSENFHGVKAVTRGQRCAIALWFLDPRH
SERDRVQADDILVRMLFSPFEMDLDSQEQMDAQGPPEPAQESUSGSEKPEDEL"
```

## Alignment Scores:

Pred. No.:	0	Length:	2524
Score:	407.00	Matches:	707
Percent Similarity:	99.16%	Conservative:	0
Best Local Similarity:	99.16%	Mismatches:	3
Query Match:	55.30%	Indels:	6
DB:	6	Gaps:	0

## ORIGIN

```
US-10-045-815-4 (1-736) x AX877139 (1-2524)
QY 1 MetAlaValArgAlaLeuLysLeuLeuThrThrLeuLeuAlaValValAlaAlaSer 20
Db 16 ATGGCGGTACGCGCTTGAAGCTGTGACCACTGCTGGCTGTGCTGGCGCTGCCCTCC 75
QY 21 GlnAlaGluValGluSerGluAlaGlyTrpGlyMetValThrProAspLeuLeuPheAla 40
Db 76 CAAGCCGAGGTTCGAGTCCGAGCAGATGGGCATGTTGACGCCCTGATCTGCTCTCC 135
QY 41 GluGlyThrAlaAlaValArgGlyAspTrpProGlyValValLeuSerMetGluArg 60
Db 136 GAGGGGACCGCAGCCTACGCGCGGGGACTGGCCCGGGTGGTCTCTGAGCATGAACGG 195
QY 61 AlaLeuArgSerArgAlaAlaLeuArgAlaLeuArgLeuArgCysArgThrClnCysAla 80
Db 196 GGGCTCGCTCCCGGCGAGCCTCCGCGCCCTTCGCTCGCTGCCGACCCAGGTGTC 255
QY 81 AlaAspPheProTrpGluLeuAspProAspTrpSerProSerProAlaGlnAlaSerGly 100
Db 256 GCGGACTTCCCGTGGAGCTGACCCCGACTGGTCTCCCGAGCCCGCCAGGCTCGGC 315
QY 101 AlaGly-AlaLeuArgAspLeuSerPhePheGlyGlyLeuLeuArgArgAlaAlaCysLe 120
Db 316 GCGGC-CGCGCTGCGCGACCTGAGCTTCTTCGGGGGCTTCTGCGCTCGCTGCTGCT 374
QY 120 uArgArgCysLeuGlyProProAlaAlaHisSerLeuSerGluGluMetGluLeuGluPh 140
Db 375 GCGCGCTGCTCGGGCGCGCGCCGACCTCGCTCAGCGAGAGATGGAGCTGGAGTT 434
QY 140 eArgLysArgSerProTyrAsnTyrLeuGlnValAlaTyrPheLysIleAsnLysLeuGl 160
Db 435 CCGCAAGCGAGCGCCCTACCACTACTGTCAGGTGCGCTACTTCAAGATCAACAAGTTGA 494
QY 160 uLysAlaValAlaAlaHisThrPhePheValGlyAsnProGluHisSerGluMetGl 180
Db 495 GAAAGCTGTGCTGAGCAGACACACCTTCTTCGTGGGCAATCTCTGAGCAGCATGGAATGA 554
QY 180 nGlnAsnLeuAspTyrTyrGlnThrMetSerGlyValLysGluAlaAspPheLysAspLe 200
Db 555 GCAGAACCTAGACTATTACCAACCATGTCTGGAGTGAAGGAGCGGACTTCAAGGATCT 614
QY 200 uGluThrGlnProHisMetGlnGluPheArgLeuGlyValArgLeuTyrSerGluGluGl 220
Db 615 TGAGACTCAACCCCATATGCAAGAATTTTCGACTGGGAGTGGCACTTACTCAGAGGAACA 674
QY 220 nProGlnGluAlaValProHisLeuGluAlaAlaLeuGlnGluTyrPheValAlaTyrGl 240
Db 675 GCACAGGAAGCTGTGCCCCACTAGAGCGCGCTGCAAGATATCTTTGTGGCTATGA 734
QY 240 uGluCysArgAlaLeuCysGluGlyProTyrAspTyrAspGlyTyrAsnTyrLeuGluTy 260
Db 735 GGAGTCCCGTGGCTCTGCGAAGGGCCCTATGACTACGATGCTACAACTACTCTTGAGTA 794
QY 260 rAsnAlaAspLeuPheGlnAlaIleThrAspHisTyrIleGlnValLeuAsnCysLysGl 280
Db 795 CAACGCTGACCTTCTCCAGGCCATCAGATCATTTACATCCAGGCTCTCACTGTAAGCA 854
QY 280 rAsnCysValThrGluLeuAlaSerHisProSerArgGluLysProPheGluAspPheLe 300
Db 855 GAACTGTGTCAGGAGCTTGTCTCCACCAGTCGAGAGAACGCTTTTGAAGACTTCTCT 914
QY 300 uProSerHisTyrAsnTyrLeuGlnPheAlaTyrTyrAsnIleGlyAsnTyrThrGlnAl 320
Db 915 CCCATCGCATTAATAATTATCTCAGTTTGCCTACTATATAACATTGGGAATTATACACAGC 974
QY 320 aGly-GluCysAlaLysThrTyrLeuLeuPhePheProAsnAspGluValMetAsnGlnA 340
Db 975 TGT-TGAATGTGCCAAGACCTATCTTCTTCTTCCCAATGACGAGGTGATGAACCAAA 1033
QY 340 snLeuAlaTyrTyrAlaAlaMetLeuGlyGluGlnHisThrArgSerIleGlyProArgG 360
|||||
```



QY 120 uArgArgCysLeuGlyProProAlaAlaHisSerLeuSerGluGluMetGluLeuGluPh 140  
Db 375 GCCTGCTGCTCGGCGCGCGCGCTCGCTCAGCGAAGATGGAGTGGAGTT 434  
QY 140 eArgLysArgSerProTyrAsnTyrLeuGlnValAlaTyrPheLysIleAsnLysLeuG 160  
Db 435 CCAGCAGCGAGCGCCCTACAACTACCTGCGAGGTGCGCTACTTCAAGATCAACAAAGTTGA 494  
QY 160 uLysAlaValAlaAlaAlaHisThrPheValGlyAsnProGluHisMetGluMetG 180  
Db 495 GAAAGCTGTTGCTGCAGCACACACCTTCTTCGTGGGCAATCTTGAGCATGGAATGCA 554  
QY 180 nGlnAsnLeuAspTyrTyrGlnThrMetSerGlyValLysGluAlaAspPheLysAspLe 200  
Db 555 GCAGAACCTAGACTATTACCAACCACTGCTGAGTGAAGAGCGGCACTTCAAGGATCT 614  
QY 200 uGluThrGlnProHisMetGlnGluPheArgLeuGlyValArgLeuTyrSerGluGluG 220  
Db 615 TGAGACTCAACCCCATATGCAAGATTTTCGACTGGGAGTGGGACTCTACTCAGAGGAACA 674  
QY 220 nProGlnGluAlaValProHisLeuGluAlaAlaLeuGlnGluTyrPheValAlaTyrG 240  
Db 675 GCCACAGAAAGCTGTGCCCACTAGAGCGCGCTGCAAGAAATACTTTGTGGCTTATGA 734  
QY 240 uGluCysArgAlaLeuCysGluGlyProTyrAsnTyrAspGlyTyrAsnTyrLeuGluTy 260  
Db 735 GAGTGTCCGTGCGCTGCGAAGAGGCGCTATGACTAGCTAGCTGCTCAACTACCTTGA 794  
QY 260 rAsnAlaAspLeuPheGlnAlaIleThrAspHisTyrIleGlnValLeuAsnCysLysG 280  
Db 795 CAACGCTGACCTTCCAGGCGCATCACAGATCATTTACATCCAGTCTCTCAACTGTAAGA 854  
QY 280 nAsnCysValThrGluLeuAlaSerHisProSerArgGluLysProPheGluAspPheLe 300  
Db 855 GAACTGTGTCCAGGAGCTTCCCTCCCAACCAAGTCGAGAGAGCCCTTTGAAGACTTCCT 914  
QY 300 uProSerHisTyrAsnTyrLeuGlnPheAlaTyrTyrAsnIleGlyAsnTyrThrGlnAl 320  
Db 915 CCATCGCATTAATATCTGCAGTTTGCCTACTATAACATTTGGGAATTATACACAGGC 974  
QY 320 aGly-GluCysAlaLysThrTyrLeuLeuPhePheProAsnAspGluValMetAsnGlnA 340  
Db 975 TGT-TGAATGTGCCAAGACCTATCTTCTTCTTCCCAATGACGAGGTGATGAACAAA 1033  
QY 340 snLeuAlaTyrTyrAlaAlaMetLeuGlyGluGluHisThrArgSerIleGlyProArgG 360  
Db 1034 ATTTGGCTATTATGCGAGCTATGCTTGGAGAAGAACACACCATTCATCCGCGCCCGTG 1093  
QY 360 luserAlaLysGluTyrArgGlnArgSerLeuLeuGluLysGluLeuLeuPhePheAlaT 380  
Db 1094 AGAGTGCAGGAGTACCGACAGCGAGCCTACTGGAAAAAGAACTGCTTTTCTCGCTT 1153  
QY 380 yrAspValPheGlyIleProPheValAspProAspSerTyrThrProGluGluValIleP 400  
Db 1154 ATGATGTTTTTGGAAATTCCTTTGTGATCCGATTCATGAGTCCAGAAAGAGTGAATC 1213  
QY 400 roLysArgLeuGlnGluLysGlnLysSerGluArgGluThrAlaValArgIleSerGlnG 420  
Db 1214 CCAGAGATTGCAAGAGAAACAGAGTCAAGCGGAAACAGCGGTACGATCTCCACAG 1273  
QY 420 luIleGlyAsnLeuMetLysGluIleGluThrLeuValGluGluLysThrLysGluSerL 440  
Db 1274 AGATTGGGAACCTTATGAAGAAATCGAGACCCCTTGTGGAGAGAGAACCAAGAGTCA 1333  
QY 440 euAspValSerArgLeuThrArgGluGlyGlyProLeuLeuTyrGluGlyIleSerLeu 460  
Db 1334 TGGATGTGAGCAGACTGACCCCGAAGGTGGCCCCCTGCTGTATGAAGGATCAGTCTCA 1393  
QY 460 hrMetAsnSerLysLeuLeuAsnGlyTyr-GlnArgValValMetAspGlyValIleSer 479  
Db 1394 CCATGAATCTCAAACTCTGAATGGTTC-CCAGCGGGTGTGTGAGCGCGGTAATCTCT 1452

QY 480 AspHisGluCysGlnGluLeuGlnArgLeuThrAsnValAlaAlaThrSerGlyAspGly 499  
Db 1453 GACCAACGAGTGTACAGGAGCTGCAGAGACTGCACCAATGTGGCAGCAACCTCAGGAGATGC 1512  
QY 500 TyrArgGlyGlnThrSerProHisThrProAsnGluLysPheTyrGlyValThrValPhe 519  
Db 1513 TACCGGGTACAGCTCCACATATCCCAATGAAAGTTCTATGGTGTCTACCTGTCTTC 1572  
QY 520 LysAlaLeuLysLeuGlyGlnGluGlyLysValProLeuGlnSerAlaHisLeuTyrTyr 539  
Db 1573 AAAGCCCTCAAGCTGGGCAAGGCAAGTTCCTCTGCAGAGTGCCACCTGTACTAC 1632  
QY 540 AsnValThrGluLysValArgArgIleMetGluSerTyrPheArgLeuAspThrProLeu 559  
Db 1633 AACGTGACGGAAGAGTGGCGCATCATGAGTCTTACTTCCGCTCGATACGCCCTC 1692  
QY 560 TyrPheSerTyrSerHisLeuValCysArgThrAlaIleGluGluValGlnAlaGluArg 579  
Db 1693 TACTTTTCT 1752  
QY 580 LysAspAspSerHisProValHisValAspAsnCysIleLeuAsnAlaGluThrLeuVal 599  
Db 1753 AAGGATGATGATCATCTCCAGTCCAGTGCAGCAACTCCCTGAAATGCCGAGACCTCTGTG 1812  
QY 600 CysValLysGluProProAlaTyrThrPheArgAspTyrSerAlaIleLeuTyrLeuAsn 619  
Db 1813 TGTGTCAAGAGCCCTCCAGCTTCCGCGACTACACCTTCCGCGACTACAGCGCCATCTTTACCTAAAT 1872  
QY 620 GlyAspPheAspGlyGlyAsnPheTyrPheThrGluLeuAspAlaLysThrValThrAla 639  
Db 1873 GGGGACTTTCGATGGCGGAACTTTTATTTCACTGAACTGGATGCCAAGACCGTGACGCA 1932  
QY 640 GluValGlnProGlnCysGlyValArgAlaValGlyPheSerSerGlyThrGluAsnProHis 659  
Db 1933 GAGGTGACGCTCAGTGTGGAAGAGCCGTGGGATCTCTTCAGGCACCTGAAACCCACAT 1992  
QY 660 GlyValLysAlaValThrArgGlyGlnArgCysAlaIleAlaLeuThrPheThrLeuAsp 679  
Db 1993 GGAGTCAAGGCTGTCCACAGGGGGCAGGCTGTGCGCATCGCCCTGTGTGTTACCTCGAC 2052  
QY 680 ProArgHisSerGluArgAspArgValGlnAlaAspAspLeuValLysMetLeuPheSer 699  
Db 2053 CTTCCAGCAGCGAGCGGACAGGTGCGAGGATGACCTGGTGAGATGCTGTGAGATGCTCTTCAGC 2112  
QY 700 ProGluGluMetAspLeuSerGlnGluGlnPro 710  
Db 2113 CCAGAAGAGATGAGCTCTCCAGGAGCAGGCC 2145  
RESULT 11  
AK027680  
LOCUS  
DEFINITION  
Homo sapiens cDNA FLJ14774 fis, clone NT2RP4000051, weakly similar  
to SYNAPTONEMAL COMPLEX PROTEIN SC65.  
ACCESSION  
AK027680  
VERSION  
AK027680.1 GI:14042539  
KEYWORDS  
cDNA capping; fis (full insert sequence).  
SOURCE  
Homo sapiens  
ORGANISM  
Homo sapiens  
REFERENCE  
1 Isogai T., Ota T., Hayashi K., Sugiyama T., Otsuki T., Suzuki Y.,  
Nishikawa T., Nagai K., Sugano S., Aotsuka S., Yoshikawa Y.,  
Matsumura H., Ishii S., Kawai Y., Saito K., Yamamoto J.,  
Kamatsuta A., Nakamura Y., Nagahara K., Masuho Y. and Sasaki N.  
NEDO human cDNA sequencing project  
2 (bases 1 to 2524)  
Isogai T. and Otsuki T.  
Direct Submission  
Submitted (10-MAY-2001) Takao Isogai, Helix Research Institute,  
Genomics Laboratory; 1532-3 Yana, Kisarazu, Chiba 292-0812, Japan  
(E-mail: genomics@hri.co.jp, Tel: 81-438-52-3975, Fax: 81-438-52-3986)

COMMENT

NEDO human cDNA sequencing project supported by Ministry of Economy, Trade and Industry of Japan; cDNA full insert sequencing: Research Association for Biotechnology; cDNA library construction, 5' - 3' - and one pass sequencing and clone selection: Helix Research Institute (supported by Japan Key Technology Center etc.) and Department of Virology, Institute of Medical Science, University of Tokyo.

## FEATURES

Location/Qualifiers

```

1. .2524
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="NT2RP400051"
/cell_line="NT2"
/cell_type="teratocarcinoma"
/clone_lib="NT2RP4"
/notes="Cloning vector: pHE18SF3-mRNA from NT2 neuronal
precursor cells after 2-weeks retinoic acid (RA)
induction."
16. .2226

```

CDS

16. .2226

```

/note="unnamed protein product"
/codon_start=1
/protein_id="BAB55291.1"
/db_xref="GI:14042540"
/translation="MAVRALKLITTLAVAAASQAQVESAGWGTTPDLLFAEGTA
AVARGDWGVLLSMERALSRAALRLRCRTCAADFPELDPDWSPPASQAGAA
ALRDLSPFGGLVLRRAACLRCLRCLPPAASLSEEMLEFRKPSYNYLVAYFKINKLE
KAAVAATLFPVNGPEHMEWQNLQDIYQTMSGWSEKLEFDLRLDPHQMFQVFLKRLVSE
NQCPQAVPHLEALQQLFYVAYBECALCEGQFYDVGNYLNVNQAELFOALTDHYIOVL
EKKQCEVPHLEASHREKQFEDELFSHNYNLQFAYINQYQAVCAKELLTFFND
EVNNQALVYAAMLGHEHSTRYSPRESAKVQRSSILKELLFYAVDFGIPVDFDGG
TLPTEVILPKRLQKQKRETRAVRLSQEIGNLMKEILFTLVEETKSLDVSRLTRSGG
PLLYEGISLTWMSKLLNGSQRVWGVVISDHECQELRLTNVAATSGDGRGQTSRHT
PNKSFYGVTFKALQDGEQKVLQSAHLVYNTVEKVRRIEMSFEDYLDLPYFSYSHL
VCTALAEVQAKDQSHGVNDVNCILNAELTAVKQCPPEATFEDYSAALYLNGDFFG
GNFYFTLEDAKTVTAEPQCGRAVFGSGVDSNENPHGVKAVTRGQRCALIAWFLDPRH
SEDRYRVQADLVNMLFSPEMDLSPQMDACQGPAPKVSGLSGSKSKPKDEL"

```

## ORIGIN

Alignment Scores:

0	Length:	2524
407.00	Matches:	707
99.16%	Conservative:	0
99.16%	Mismatches:	3
55.30%	Indels:	6
9	Gaps:	0
DB:		

US-10-045-815-4 (1-736) x AK027680 (1-2524)

1	MetAlaValArgAlaLeuLysLeuThrThrLeuLeuAlaValAlaAlaSer	20	Oy
16	ATGGCGGTACGCGGCTTGAAGCTCTGACCACACTGCTGGCTCTCGTGGCGCTGCTCC	75	Db
21	GlnAlaGluValGluSerGluAlaGlyTrpGlyMetValThrProAspLeuLeuPheAla	40	Oy
76	CAAGCCGAGGTCTGAGTCCGAGGCAAGATGGGCATGGTGACGCTGATCTGCTTCGGC	135	Db
41	GluGlyThrAlaAlaTyrAlaArgGlyAspTrpProGlyValValLeuSerMetGluArg	60	Oy
136	GAGGGACCGCAGGCTACGGCGCGGGAGATGGCCGGGTGCTCTGACATGGGAACG	195	Db
61	AlaLeuArgSerArgAlaAlaLeuArgAlaLeuArgLeuArgCysArgThrGlnCysAla	80	Oy
196	CGCTGCGCTCCCGGCGAGCCCTCCGCGCCTTCGCTGCGTCCGACCCAGGTGTC	255	Db
81	AlaAspPheProTrpGluLeuAspProAspTrpSerProSerProAlaGlnAlaSerGly	100	Oy
256	GCCGACTTCCGTGGGAGCTGACCCCGATGTGTCCCCAGCCGCGCCAGGCTCTGGGC	315	Db
101	AlaGly-AlaLeuArgAspLeuSerPhePheGlyGlyLeuLeuArgArgAlaAlaCysLe	120	Oy
316	GCQCG-CGCGCTCGCGACCTGAGCTTCTTCGGGGGCTTCTCGGTCGGGGCTGCTGCT	374	Db

```
Db 1453 GACACGAGTGTGAGGAGTGCAGAGATGACCAATGTGCGCAACCTCAGAGATGCC 1512
Qy 500 TTAAGGAGTGTGAGGAGTGCAGAGATGACCAATGTGCGCAACCTCAGAGATGCC 1512
Db 1513 TACCGGGTTCAGACCTCCCAATATCTCCCAATGAAAGTTCATGCTGCTCCTCTC 1572
Qy 520 LysAlaLeuLysLeuGlyGlnGluGlyLysValProLeuGluSerAlaHisLeuTyr 539
Db 1573 AAAGCCCTCAAGCTGGCGGCAAGAAGCAAAAGTTCCTCTGCGAGAGTCCCACTGACTAC 1632
Qy 540 AsnValThrGluLysValArgArgGileMetGluSerTyrPheArgLeuAspThrProLeu 559
Db 1633 AAGCTGACGAGAGAGTGGCGGCATCATGAGTCTCTACTTCGCGCTGGATAGCCCTC 1692
Qy 560 TyrPheSerTyrSerHisLeuValCysArgThrAlaIleGluGluValGlnAlaGluArg 579
Db 1693 TACTTTCTTACTCTCATCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1752
Qy 580 LysAspAspSerHisProValHisValAspAsnCysIleLeuAsnAlaGluThrLeuVal 599
Db 1753 AAGGATGATAGTATCCAGTCCAGTGGCAACCTGATCTGATGCGGAGACCTCGTG 1812
Qy 600 CysValLysGluProProAlaTyrThrPheArgAspTyrSerAlaIleLeuTyrLeuAsn 619
Db 1813 TGTGTCAAAGAGGCCCCAGGCTACACCTTCGCGGACTACAGCGCCATCTTTACCTAAT 1872
Qy 620 GlyAspPheAspGlyGlyAsnPheTyrPheThrGluLeuAspAlaLysThrValThAla 639
Db 1873 GGGGACTTCGATGCGGAACTTTTATTCACCTGAACCTGATGCAAGACCTGAGCGCA 1932
Qy 640 GluValGlnProGlnCysGlyArgAlaValGlyPheSerSerGlyThrGluAsnProHis 659
Db 1933 GAGGTGAGGCTGCTGAGGAGGCGTGGGATCTCTTCGAGCTGAAACCCACAT 1992
Qy 660 GlyValLysAlaValThrArgGlyGlnArgCysAlaIleAlaLeuTyrPheThrLeuAsp 679
Db 1993 GGAGTGAGGCTGTCACCGGGGCGAGCGTGTGCTGCTGCTGCTGCTGCTGCTGCTG 2052
Qy 680 ProArgHisSerGluArgAspArgValGlnAlaAspAspLeuValLysMetLeuPheSer 699
Db 2053 CCTGCACACGAGCGGAGCGGAGTGCAGGAGATGACCTGTTGAGATGCTCTTCAGC 2112
Qy 700 ProGluGluMetAspLeuSerGlnGluGlnPro 710
Db 2113 CCAGAGAGATGAGCTCTCCAGGAGCAGCCC 2145

RESULT 12
AF097431 2829 bp mRNA linear PRI 22-NOV-2000
LOCUS Homo sapiens GROS1-S protein mRNA, complete cds.
ACCESSION AF097431
VERSION AF097431.1 GI:11127635
KEYWORDS
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 2829)
Kaul, S.C., Sugihara, T., Yoshida, A., Nomura, H. and Wadhwa, R.
Gros1, a potential growth suppressor on chromosome 1: its identity
to basement membrane-associated proteoglycan, leprecan
Oncogene 19 (32), 3576-3583 (2000)
JOURNAL 2040537
MEDLINE
PUBMED 10951563
REFERENCE 2 (bases 1 to 2829)
AUTHORS Wadhwa, R., Sugihara, T. and Kaul, S.
TITLE Direct Submission
JOURNAL Submitted (06-OCT-1998) Group 4, Chugai Research Institute of
Molecular medicine, 153-2 Nagai, Niihari, Ibaraki 300-41, Japan
FEATURES
Location/Qualifiers
1..2829
/organism="Homo sapiens"
```

```
/mol_type="mRNA"
/db_xref="taxon:9606"
/chromosome="1"
/tissue_type="testis"
52..1143
/notes="member of protein B family"
/codon_start=1
/product="GROS1-S protein"
/protein_id="AAG31018.1"
/db_xref="GI:11127636"
/translation="MAVRALKLTLTLAVAAASQAESEAGWMTPTDLLFAEGTA
AVAGDPGVVLSMERALRLRCRTOCAADFPEWLPDWSPPSPAQASGAG
ALRDLSPFGILLRRAACLRRCGLPPAAHLSSEMELEFRKSPNYLQVAFKINKLE
KAVAAATFPVGNPEHMOQNLQYCTMSGVKEADFKOLETQPHMVEFRGLVLYSE
EQGAVFELSRALQEIFVAYEBCALCEGPDYDGNLYEYNADLFOALTDHYLOVL
NCKQNVTELSHSPSRKPFDFLPESHYNYLQFAYINIGNITQAGECAKTYLLFPND
EVMNQNLAYAAMLGEHTRSIGPREQGT"

ORIGIN
Alignment Scores:
Pred. No.: 0 Length: 2829
Score: 360.00 Matches: 360
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 48.91% Indels: 0
DB: 9 Gaps: 0

US-10-045-815-4 (1-736) x AF097431 (1-2829)
Qy 1 MetAlaValArgAlaLeuLysLeuLeuThrThrLeuLeuAlaValAlaAlaLaser 20
Db 52 ATGGCGGTACGGCGCTTGAAGCTGTGACACACATGCTGGTGTCTGTCGCGCTGCCCTCC 111
Qy 21 GlnAlaGluValGluSerGluAlaGlyTTPGlyMetValThrProAspLeuPheAla 40
Db 112 CAAGCGAGGTGAGTCCGAGCAGGATGGGCGATGGTGGCGCTGATCTCTCTCGCC 171
Qy 41 GluGlyThrAlaAlaTyrAlaArgGlyAspTTPProGlyValValLeuSerMetGluArg 60
Db 172 GAGGGGACCGGAGCTTACGGCGCGGGGACTGGCCCGGGTGTCTCTGAGCATGACCG 231
Qy 61 AlaLeuArgSerArgAlaAlaLeuArgAlaLeuArgLeuArgCysArgThrGlnCysAla 80
Db 232 GCGCTGCGCTCCCGGCGAGCCCTCCGCGCTTCCTGCGTGGCGCAGCCAGTGTGCC 291
Qy 81 AlaAspPheProTTPGluLeuAspProAspTTPSerProAlaGlnAlaSerGly 100
Db 292 GCGGACTTCCCGTGGAGCTGAGCCCGGACTGGTCCCGCGCGCGCGCGCTCGGCG 351
Qy 101 AlaGlyAlaLeuArgAspLeuSerPhePheGlyGlyLeuLeuArgArgAlaAlaCysLeu 120
Db 352 GCGGCGCGCTCGCGGAGCTGAGCTTCTCGGGGCGCTTCTGCGTGGCGCTGCCCTGCC 411
Qy 121 ArgArgCysLeuGlyProProAlaAlaHisSerLeuSerGluMetGluLeuGluPhe 140
Db 412 CCGCGCTGCTCGGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 471
Qy 141 ArgLysArgSerProTyrAsnTyrLeuGlnValAlaTyrPheLysIleAsnLysLeuGlu 160
Db 472 CGCAAGCGGAGCGCGCTACAACTACCTGCGAGTTCGCTTCTCAAGATCAACAAGTTGGAG 531
Qy 161 LysAlaValAlaAlaAlaHisThrPhePheValGlyAsnProGluHisMetGluMetGln 180
Db 532 AAAGCTGTTGCTGCGAGCACACACCTTCTTCGTTGGGCAATCTCCGAGCACATGAATGCGAG 591
Qy 181 GlnAsnLeuAspTyrTyrGlnThrMetSerGlyValLysGluAlaAspPheLysAspLeu 200
Db 592 CAGAAGCTGAGATATTACCAACCATGCTGAGTGAAGAGGCGCGCGCTTCAAGGATCTT 651
Qy 201 GluThrGlnProHisMetGlnGluPheArgLeuGlyValArgLeuTyrSerGluGln 220
Db 652 GAGACTCAACCCCATATGCAAGAAATTCGACTGGGAGTGCAGTCTTACTCTCAGAGAACAG 711
```



QY 221 ProGlnGluAlaValProHisLeuGluAlaAlaLeuGlnGluTyrPheValAlaTyrGlu 240  
 DB 712 CCACAGGAAGCTGTGCCCCACCCTAGAGGCGCGCTGCAAGAAATCTTTGTGGCCTATGAG 771  
 QY 241 GluCysArgAlaLeuGluGluGlyProTyrAspTyrAspGlyTyrAsnTyrLeuGluTyr 260  
 DB 772 GAGTCCCGTGGCCCTCTGGAAGGCGCCCTATGACATGATGCTCACTACCTTTGAGTAC 831  
 QY 261 AsnAlaAspLeuPheGlnAlaIleThrAspHisTyrIleGlnValLeuAsnCysLysGln 280  
 DB 832 AACGCTGACCTCTTCCAGGCCATCACACATCATATCATCCAGTCTCTCACTGTAAGCAG 891  
 QY 281 AsnCysValThrGluLeuAlaSerHisProSerArgGluLysProPheGluAspPheLeu 300  
 DB 892 AACTGTGTACGGAGCTTCTTCCACCACCAAGTGCAGAGAGCCCTTTGAGACTTCTCCTC 951  
 QY 301 ProSerHisTyrAsnTyrLeuGlnPheAlaTyrTyrAsnIleGlyAsnTyrThrGlnAla 320  
 DB 952 CCATCGCATTAATATCTGCAGTTTCCCTACTATACATTTGGAAATATACAAAGCT 1011  
 QY 321 GlyGluCysAlaLysThrTyrLeuPhePheProAsnAspGluValMetAsnGlnAsn 340  
 DB 1012 GGTGAATGTGCAGACCTATCTCTCTTCCCAATGACGAGGTGATGAACCAAAAT 1071  
 QY 341 LeuAlaTyrTyrAlaAlaMetLeuGluGluHisThrArgSerIleGlyProArgGlu 360  
 DB 1072 TTGGCTATTATGACGTATGCTTGGAGAGAACACACACAGATCATCGGCCCCCGTGAG 1131  
 RESULT 13  
 BC015309 1562 bp mRNA linear PRI 30-SEP-2003  
 Locus  
 DEFINITION Homo sapiens leucine proline-enriched proteoglycan (leprecan) 1,  
 mRNA (CDNA clone IMAGE:4395073), partial cds.  
 ACCESSION BC015309  
 VERSION BC015309.2 GI:34782974  
 KEYWORDS  
 SOURCE Homo sapiens (human)  
 ORGANISM  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.  
 REFERENCE 1 (bases 1 to 1562)  
 AUTHORS Strausberg, R.L., Feingold, E.A., Grouse, L.H., Derge, J.G.,  
 Klaus, R.D., Collins, F.S., Wagner, L., Shenmen, C.M., Schuler, G.D.,  
 Altschul, S.F., Zeeberg, B., Buetow, K.H., Schaefer, C.F., Bhat, N.K.,  
 Hopkins, R.F., Jordan, H., Moore, T., Max, S.I., Wang, J., Hsieh, F.,  
 Diatchenko, L., Narusina, K., Farmer, A.A., Rubin, G.M., Hong, L.,  
 Stapleton, M., Soares, M.B., Bonaldo, M.F., Casavant, T.L.,  
 Scheetz, T.E., Brownstein, M.J., Usdin, T.B., Toshiyuki, S.,  
 Carninci, P., Prange, C., Raha, S.S., Loquellano, N.A., Peters, G.J.,  
 Abramson, R.D., Mullahy, S.J., Bosak, S.A., McEwan, P.J.,  
 McKernan, K.J., Malek, J.A., Gunaratne, P.H., Richards, S.,  
 Worley, K.C., Hale, S., Garcia, A.M., Gay, L.J., Hulyk, S.W.,  
 Villalón, D.K., Muzny, D.M., Sodergren, E.J., Lu, X., Gibbs, R.A.,  
 Fahey, J., Helton, E., Kettman, M., Madan, A., Rodriguez, S.,  
 Sanchez, A., Whitting, M., Madan, A., Young, A.C., Shevchenko, Y.,  
 Bouffard, G.G., Blakesley, R.W., Touchman, J.W., Green, E.D.,  
 Dickson, M.C., Rodriguez, A.C., Grimwood, J., Schmutz, J., Myers, R.M.,  
 Butterfield, Y.S., Krzywinski, M.I., Skalska, U., Smalios, D.E.,  
 Schnerch, A., Schein, J.E., Jones, S.J. and Marra, M.A.  
 Generation and initial analysis of more than 15,000 full-length  
 human and mouse cDNA sequences  
 Proc. Natl. Acad. Sci. U.S.A. 99 (26), 16899-16903 (2002)  
 REFERENCE 2 (bases 1 to 1562)  
 JOURNAL Strausberg, R.  
 MEDLINE Direct Submission  
 PUBMED Submitted (01-OCT-2001) National Institutes of Health, Mammalian  
 TITLE Gene Collection (MGC), Cancer Genomics Office, National Cancer  
 JOURNAL Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,  
 USA  
 REMARK NIH-MGC Project URL: <http://mgc.nci.nih.gov>  
 COMMENT On Sep 16, 2003 this sequence version replaced gi:15929775.

Contact: MGC help desk  
 Email: [cgabbs-r@mail.nih.gov](mailto:cgabbs-r@mail.nih.gov)  
 Tissue Procurement: ATCC  
 cDNA Library Preparation: Life Technologies, Inc.  
 cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)  
 DNA Sequencing by: Sequencing Group at the Stanford Human Genome  
 Center, Stanford University School of Medicine, Stanford, CA 94305  
 Web site: <http://www.shgc.stanford.edu>  
 Contact: (Dickson, Mark) [mcd@paxil.stanford.edu](mailto:mcd@paxil.stanford.edu)  
 Dickson, M., Schmutz, J., Grimwood, J., Rodriguez, A., and Myers, R. M.

Clone distribution: MGC clone distribution information can be found  
 through the I.M.A.G.E. Consortium/LLNL at: <http://image.llnl.gov>  
 Series: IRAC Place: 27 Row: c Column: 1.

FEATURES  
 Location/Qualifiers  
 1..1562  
 /organism="Homo sapiens"  
 /mol\_type="mRNA"  
 /db\_xref="taxon:9606"  
 /clone="IMAGE:4395073"  
 /tissue\_type="Duodenum, adenocarcinoma"  
 /clone\_lib="NIH MGC\_88"  
 /lab\_host="DH10B"  
 /note="Vector: pCMV-SPORT6"  
 <1..1562  
 /gene="LEPRE1"  
 /note="synonym: GROS1"  
 /db\_xref="LocusID:64175"  
 <1..1226  
 /gene="LEPRE1"  
 /codon\_start=3  
 /product="LEPRE1 protein"  
 /protein\_id="AAH15309.2"  
 /db\_xref="GI:34782975"  
 /db\_xref="LocusID:64175"  
 /translation="DPRVREVNQNLAYAAMLGEHTSTIGPRESAKYRQRLLEK  
 ELLFPAYDFGIPVDPSWTPEVVKRLQEKSKSETRAVRISQIGLNMKEIETL  
 VEETKESLDGVRSLTREGGPLYEGISLTMNSKLNGSQRVVMGVISDHQCEQLQRL  
 INVAATSGDYRGSTGSPHTNEKFYGVTVFKALKGQEGKVPLOSAHLIYVNTSEKVR  
 IMESYRLDTPLFYSLVCAVTAIEVQAERKDDSHPVHVDNCILNAETLVCVKEPP  
 AYTFRDYGAILYLGDFDGGNFYETELDAKTATVAEQPCQGRVGFSGSTENPHGVA  
 VTGRQCAIALWFTLDPHRSRDRVQADLLVKMLFSPEMDLSQEQPLDAQCGPPEA  
 QESLGSQSKPKDEL"  
 426..1046  
 /gene="LEPRE1"  
 /note="P4HC; Region: Prolyl 4-hydroxylase alpha subunit  
 homologues. Mammalian enzymes catalyse hydroxylation of  
 collagen, for example. Prokaryotic enzymes might catalyse  
 hydroxylation of antibiotic peptides. These are  
 2-oxoglutarate-dependent dioxygenases, requiring  
 2-oxoglutarate and dioxygen as cosubstrates and ferrous  
 iron as a cofactor"  
 /db\_xref="CDD:smart00702"

misc\_feature  
 Alignment Scores:  
 Pred. No.: 0 Length: 1562  
 Score: 301.00 Matches: 401  
 Percent Similarity: 99.50% Conservatve: 0  
 Best Local Similarity: 99.50% Mismatches: 1  
 Query Match: 40.90% Indels: 2  
 DB: 9 Gaps: 0  
 US-10-045-815-4 (1-736) x BC015309 (1-1562)  
 QY 335 GluValMetAsnGlnAsnLeuAlaTyrTyrAlaAlaMetLeuGlyGluGluHisThrArg 354  
 DB 18 GAGGTGATGACCAACAAATTTGGCCTATTATGACATGCTTGGAGAGAACACACACAGA 77  
 QY 355 SerIleGlyProArgGluSerAlaLysGluTyrArgGlnArgSerLeuLeuGluLysGlu 374  
 DB 78 TCCATCGGCCCCCGTGAGAGTGCCAGAGGTACCGACAGCGAGCCTACTCGAAAAAGAA 137



```
QY 375 LeuLeuPhePheAlaTyAspValPheGlyIleProPheValAspProAspSerTyrThr 394
Db 138 CTGCTTTCTCGCTTATGATGTTTGGAAATTCCTTTGTGGATCCGGATTCATGCACT 197

QY 395 ProGluValIleProLysArgLeuGlnGluLysGlnLysSerGluArgGluThrAla 414
Db 198 CCAGAAAGATGATTCCTCAAGATTTGCAAGAGAAACAGAAAGTCAGACGGGAAACAGCC 257

QY 415 ValArgIleSerGlnGluIleGlyAsnLeuMetLysGluIleGluThrLeuValGluGlu 434
Db 258 GTACGCATCTCCAGGAGATTGGGAACCTTATGAGGAAATCAGACCTTGTGGAAGAG 317

QY 435 LysThrLysGluSerLeuAspValSerArgLeuThrArgGluGlyProLeuLeuTyr 454
Db 318 AAGACCAAGAGTCACTGGATGTGAGCAGACTACCCGGGAAGTGGCCCCCTGCTGTAT 377

QY 455 GluGlyIleSerLeuThrMetAsnSerLysLeuLeuAsnGlyTyr-GlnArgValValMe 474
Db 378 GAAGGCATGATCTCACCAGACTCCAACTCTGATGTTTC-CCAGCGGTGTGTAT 436

QY 474 tAspGlyValIleSerAspHisGluCysGlnGluLeuGlnArgLeuThrAsnValAla 494
Db 437 GGACGGCGTAACTCTGACCACGAGTGTGAGGAGCTGACAGACTGACCAATGTGCAGC 496

QY 494 aThrSerGlyAspGlyTyrArgGlyGlnThrSerProHisThrProAsnGluLysPheTy 514
Db 497 AACCTCAGGAGATGGCTACCGGGTCCAGACTCCCACTACTCCCAATGAAAGATCTTA 556

QY 514 rGlyValThrValPheLysAlaLeuLysLeuGlyGlnGluLysValProLeuGlnSe 534
Db 557 TGGTGTCACTGTCTTCAAGCCCTCAAGCTGGGCAAGGCAAGTTCCTCTGCAGAG 616

QY 534 rAlaHisLeuTyrTyrAsnValThrGluLysValArgArgIleMetGluSerTyrPheAr 554
Db 617 TGCCCACTGTACTACAACGTGACGGAAGAGTGGCGCGCATCATGGAGTCTTACCTCCG 676

QY 554 gLeuAspThrProLeuTyrPheSerTyrSerHisLeuValCysArgThrAlaIleGlu 574
Db 577 CTGGATACGCCCTCTACTTTTCTACTCTCATCTGGTGTGGCGCACTGCCATCGAGA 736

QY 574 uValGlnAlaGluArgLysAspSerHisProValHisValAspAsnCysIleLeuAs 594
Db 737 GGTCCAGCGCAGAGGAAGATGATGATCATCCAGTCCACGTGGCAAACTGCATCTCTGAA 796

QY 594 nAlaGluThrLeuValCysValLysGluProProAlaTyrPheArgAspTyrSerAl 614
Db 797 TGGCGAGACCTCTGTGTGTCTAAAGAGCCCCAGCCCTACACCTTCGCGACTACAGCGC 856

QY 614 aIleLeuTyrLeuAsnGlyAspPheAspGlyGlyAsnPheTyrPheThrGluLeuAspAl 634
Db 857 CATCTTTTACTTAATCGGGACTTCGATGGCGGAACCTTTTATTTCACTGAATGGATGC 916

QY 634 aLysThrValThrAlaGluValGlnProGlnCysGlyArgAlaValGlyPheSerSerG 654
Db 917 CAAGACCGTCGCGCAGAGTGCAGCCCTCAGTGTGGAAAGCGCGTGGGATTCCTTCAGG 976

QY 654 YThrGluAsnProHisGlyValLysAlaValThrArgGlyGlnArgCysAlaIleAlaLe 674
Db 977 CACTGAAAACCCATCGAGTGAAGGTGTTCACAGGGGCGAGCGCTGTGCTGCTCGCCCT 1036

QY 674 uTTPPheThrLeuAspProArgHisSerGluUArgAspArgValGlnAlaAspAspLeuVa 694
Db 1037 GTGGTTTCACTGACCTCGACACAGCGAGCGGAGCAGGGTGCAGGCAGATGACCTGGT 1096

QY 694 lLysMetLeuPheSerProGluGluMetAspLeuSerGlnGluGlnProLeuAspAlaGl 714
Db 1097 GAAGATGCTCTTACGCCACAGAAGAGATGACCTCTCCAGAGAGCAGCCCTGTGATGCCCA 1156

QY 714 nGlnGlyProProGluProAlaGlnGluSerLeuSerGlySerGluSerLysProLysAs 734
Db 1157 GCAGGGCCCCCGAACCTGACAAAGAGTCTCTCTCAGGCAGTGAATCGAAGGCCCAAGGA 1216
```

```
QY 734 pGluLeu 736
Db 1217 TCAGCTA 1223

RESULT 14
AX880254
LOCUS AX880254
DEFINITION Sequence 15159 from Patent EP1074617.
ACCESSION AX880254
VERSION AX880254.1 GI:40034990
KEYWORDS Homo sapiens (human)
SOURCE Homo sapiens
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1
AUTHORS Ota,T., Isogai,T., Nishikawa,T., Hayashi,K., Saito,K., Yamamoto,J.,
Ishii,S., Sugiyama,T., Wakamatsu,A., Nagai,K. and Otsuki,T.
Primers for synthesising full-length cDNA and their use
Patent: EP 1074617-A 15159 07-FEB-2001;
Research Association for Biotechnology (JP)
FEATURES
Location/Qualifiers
1..2993
/organism="Homo sapiens"
/mol_type="unassigned DNA"
/db_xref="taxon:9606"
42..2456
/note="unnamed protein product"
/codon_start=1
/protein_id="CAE90869.1"
/db_xref="GI:40034991"
/translation="MAVRALKLTLTLLAVAAASAEVESEAGMGWTPDILLFAEGTA
AVRGDWPVLSMERALSRAALRLRCRTCAADFPWELDPDWSPSPAQASGAA
ALRDLSPGGLLRRAACLRCLGPPAAHLSSEMELEPRKSPYNLYOVAYEKLNKLE
KAVAAHTFFVGNPEHMEQNLDTYTMSGVKEADFKDLTQPHMQBFRGLVLYSE
EQQAVPHEALAOEYVAYEERCALEGYDYGVNLYEYNADLFQAITDHYIQVL
NCKMNCVELASHPSREKPFDFLPSHNYLQFAYINIGNYTQACEKATLLFFPND
EVMKNCILAYAAALGMEHRETRISQISGRSAKEYSRSLLEKELFFAYDVGIPFVDPDS
WTPGEVTPKRLQEKQKSERETAVRISQISGRSAKEYSRSLLEKELFFAYDVGIPFVDPDS
PLLYGISLTWNSKLLNGSQRVAVDVTSDHECQELQRLTNVAATSGDVGRTGSPHT
ENRKYGVTVFKALKGEGKVPQSAHLYNVTEKVRIMESFRDLPFLYFSYSHL
VCRTAIEVQAEKDDSHPVHNCILNAETFLVCVKPPATFRDYISAILYNGDFDG
GNFTYELDAKTVAETVQPCQGRVAFSGFSTENPHGVKAVTRGQRCALAFITLDPRH
SERVRAARAGSSWCCGDFPFPWFAPLFPKSHCOMLHERHSTWDTSSNALSWSHC
LVLPPAVNGIQVGKEVNTGSDABFLVPSLPTSAVLQRVGPAKGEMSLGPLRLNLC
PLGSSS"

Alignment Scores:
Pred. No.: 2,55e-314 Length: 2993
Score: 295.00 Matches: 681
Percent Similarity: 98.84% Conservative: 0
Best Local Similarity: 98.84% Mismatches: 4
Query Match: 40.08% Indels: 8
DB: 6 Gaps: 0

US-10-045-815-4 (1-736) x AX880254 (1-2993)

QY 1 MetAlaValArgAlaLeuLysLeuLeuThrThrLeuLeuAlaValAlaAlaAlaSer 20
Db 42 ATGGCGGTACGCGCTTGAAGCTGCTGACCACTGCTGGCTGCTGCGCGCTGCCCTCC 101

QY 21 GlnAlaGluValGluSerGluAlaGlyTrpGlyMetValThrProAspLeuLeuPheAla 40
Db 102 CAAGCGGAGTTCGAGTCCGAGCAGGATGGGCGCATGTGACGCCCTGATCTGCTTCGCC 161

QY 41 GluGlyThrAlaAlaTyrAlaArgGlyAspTyrProGlyValValLeuSerMetGluArg 60
Db 162 GAGGGGACCCGAGCCCTACCGCGCGGGGACTGGCCCGGGTGGTCTTGAAGCATGGAACGG 221

QY 61 AlaLeuArgSerArgAlaAlaLeuArgAlaLeuArgLeuArgCysArgThrGlnCysAla 80
Db 222 GCGCTGCGCTCCCGGGGAGCCCTTCGCGCCCTTCGCTGCGCTGCGCACCCAGTGTGCC 281
```

QY 81 AlaAspPheProTyrGluLeuAspProAspTyrSerProSerProAlaGlnAlaSerGly 100  
DB 282 GCCGACTTCCTCGTGGGAGCTGGACCCCGAGCTGGTCCCCCGCCGCGCCAGGCTCGGGC 341  
QY 101 AlaGly-AlaLeuArgAspLeuSerPhePheGlyGlyLeuLeuArgAlaAlaCysLe 120  
DB 342 GCCGC-CGCCCTGGCGGACCTGAGCTTCCTTCGGGGGCTTCCTCGCTCGCGCTGCTGCCT 400  
QY 120 uArgArgCysLeuGlyProProAlaAlaHisSerLeuSerGluGluMetGluLeuGluPh 140  
DB 401 CGCGCGCTTCGGCG 460  
QY 140 eArgGlyArgSerProTyrAsnTyrLeuGlnValAlaTyrPheLysIleAsnLysLeuG 160  
DB 461 CGCAAGCGGAGCCCTACAACTACCTGCGAGTGGCTACTTCAAGATCAACAAGTTGGA 520  
QY 160 uLysAlaValAlaAlaAlaHisThrPheValGlyAsnProGluHisMetGluMetG 180  
DB 521 GAAAGCTGTTCGTCGAGCACACACTTCTTCGTCGGCAATCCTGAGCACATGGAATGCA 580  
QY 180 nGlnAsnLeuAspTyrTyrGlnThrMetSerGlyValLysGluAlaAspPheLysAspLe 200  
DB 581 GCAGAACCTAGACTATTACCAAAACCATGCTGAGAGTCAAGAGCGCGACTTCAAGGATCT 640  
QY 200 uGluThrGlnProHisMetGlnGluPheArgLeuGlyValArgLeuTyrSerGluGluG 220  
DB 641 TGAGACTCAACCCCATATGCAAGAAATTCGACTGGGAGTGGACTCTACTCAGAGGAACA 700  
QY 220 nProGlnGluAlaValProHisLeuGluAlaAlaLeuGlnGluTyrPheValAlaTyrG 240  
DB 701 CCCACAGGAAGCTGTGCCACCTAGAGGCGCGCTGCAAGATACATTTTGTGGCTATGA 760  
QY 240 uGluCysArgAlaLeuCysGluGlyProTyrAspTyrAspGlyTyrAsnTyrLeuGlu 260  
DB 761 GGAGTGGCGTCCCTCTCGCAAGGCGCTATGACTAGCATGGCTTACAACTACCTTGAGTA 820  
QY 260 rAsnAlaAspLeuPheGlnAlaThrAspHisTyrIleGlnValLeuAsnCysLysG 280  
DB 821 CAACGCTGACCTCTTCAGGCGCATCAGAGATCATATCATCAGGTCTCAACTGTAGCA 880  
QY 280 nAsnCysValThrGluLeuAlaSerHisProSerArgGluLeuProPheGluAspPheLe 300  
DB 881 GAACCTGTGTCAGGAGCTGCTCCCAACCAAGTCGAGAGAGCCCTTTTGAAGACTTCCT 940  
QY 300 uProSerHisTyrAsnTyrLeuGlnPheAlaTyrTyrAsnIleGlyAsnTyrThrGlnAl 320  
DB 941 CCCATCGCATTAATATATCTGCAAGTTGCTTACTATAAATTGGGAATTTATACAGGC 1000  
QY 320 aGly-GluCysAlaLysThrTyrLeuLeuPhePheProAsnAspGluValMetAsnGln 340  
DB 1001 TGT-TGAATGTGCCAAGACCTATCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCT 1059  
QY 340 snLeuAlaTyrTyrAlaAlaMetLeuGlyGlnGluHisThrArgSerIleGlyProArg 360  
DB 1060 ATTTGGGCTATTATGCACTATGCTTGGAGAAGAACACACAGATCATCCGCCCCCGTG 1119  
QY 360 luSerAlaLysGluTyrArgGlnArgSerLeuLeuGluLysGluLeuLeuPhePheAla 380  
DB 1120 AGATGCCAAGAGTACCGACAGGAGCTTACTGGAAAAGAACTGCTTTTCTTCTGCTT 1179  
QY 380 yrAspValPheGlyIlePhePheValAspProAspSerTyrThrProGlu-GluValIle 399  
DB 1180 ATGATGTTTTTGGAAATTCCTTTTGTGATCCGATTCATGAGCTCCAGG-AGAAGTGATT 1238  
QY 400 ProLysArgLeuGlnGluLysGlnLysSerGluArgGluThrAlaValArgIleSerGln 419  
DB 1239 CCCAAGAGATTGCAAGAGAAACAGAACTCAGACCGGAAACAGCGCTTACCATCTCCAG 1298  
QY 420 GluIleGlyAsnLeuMetLysGluIleGluThrLeuValGluGluLysThrLysGluSer 439  
DB 1299 GAGATTGGGAACCTTATGAAGGAATCGAGACCTTGTGGAAGAGAAGCAAGAGGAGTCA 1358

QY 440 LeuAspValSerArgLeuThrArgGluGlyProLeuLeuTyrGluGlyIleSerLeu 459  
DB 1359 CTGGATGTGAGCAGACTGACCGGGAAGTGGCCCCCTGCTGTATGAAGGCATCAGTCTC 1418  
QY 460 ThrMetAsnSerLysLeuLeuAsnGlyTyr-GlnArgValValMetAspGlyValIleSe 479  
DB 1419 ACCATGAACCTCAAACTCCTGAATGGTTC-CAGCGGGTGGTGGTGGACGGCGTAATCTC 1477  
QY 479 rAspHisGluCysGlnGluLeuGlnArgLeuThrAsnValAlaAlaThrSerGlyAspG 499  
DB 1478 TGACCAACGAGTCTCAGGAGCTGCAGAGACTGCACATGTGGCAGCAACCTCAGAGATGG 1537  
QY 499 yTyrArgGlyGlnThrSerProHisThrProAsnGluLysPheTyrGlyValThrValPh 519  
DB 1538 CTACCGGGTTCAGACTCCCACTATCCCAATGAAAAGTTCTATGGTGTGCTGCTGCTT 1597  
QY 519 eLysAlaLeuLysLeuGlyGlnGluGlyLysValProLeuGlnSerAlaHisLeuTyrTy 539  
DB 1598 CAAGCCCTCAAGCTGGGCAAGAGGCAAGTTCCTCTGCAGAGTGGCCACCTGTACTA 1657  
QY 539 rAsnValThrGluLysValArgArgIleMetGluSerTyrPheArgLeuAspThrProLe 559  
DB 1658 CAACGTGACGGAAGAGTGGCGGCATCATGAGTCTTCCGCTCGATACGCCCT 1717  
QY 559 uTyrPheSerTyrSerHisLeuValCysArgThrAlaIleGluGluValGlnAlaGluAr 579  
DB 1718 CTACTTTTCTTACTCTCATCTGTGTGGCGCACTGCCATCGAAGAGGTCCAGGCAAGAG 1777  
QY 579 glyAspAspSerHisProValHisValAspAsnCysIleLeuAsnAlaGluThrLeuVa 599  
DB 1778 GAAGCATGATAGTATCCAGTCCAGTGCAGCAACTGCATCTGAAATCCGAGACCTCGT 1837  
QY 599 lCysValLysGluProProAlaTyrThrPheArgAspTyrSerAlaIleLeuTyrLeuAs 619  
DB 1838 GTGTGTCAAAAGAGCCCGCAGCTACACCTTCGGGACTACAGCGCCATCTTTTACCTAAA 1897  
QY 619 nGlyAspPheAspGlyGlyAsnPheTyrPheThrGluLeuAspAlaLysThrValThrAl 639  
DB 1898 TGGGCACTTCGATGGCGGAACCTTTTATTTCATGAACTGGATGCCAAGACCGTGAGCGC 1957  
QY 639 aGluValGlnProGlnCysGlyArgAlaValAlaGlyPheSerSerGlyThrGluAsnProHi 659  
DB 1958 AGAGTGCAGCCCTCAGTGTGGAAGAGCGGTGGGATTTCTTTTCAGGCACTGAAACCCACA 2017  
QY 659 sGlyValLysAlaValThrArgGlyGlnArgCysAlaIleAlaLeuTyrPheThrLeuAs 679  
DB 2018 TGGAGTGAAGCTGTCCAGGCGGCGAGCTGTGCCATCGCCCTGTGTTTCACCTCGGA 2077  
QY 679 pProArgHisSerGluArg 695  
DB 2078 CCTCGACACAGCAGCGCG 2096  
RESULT 15  
BD158280 2993 bp DNA linear PAT 17-JAN-2003  
LOCUS Primer for synthesizing full-length cDNA and use thereof.  
DEFINITION BD158280  
ACCESSION BD158280  
VERSION BD158280.1 GI:27864038  
KEYWORDS JP 2002191363-A/13123.  
SOURCE Homo sapiens (human)  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
REFERENCE 1 (bases 1 to 2993)  
AUTHORS Ota,T., Isogai,T., Nishikawa,T., Hayashi,K., Saito,K., Yamamoto,J.,  
Ishii,S., Sugiyama,T., Wakamatsu,A., Nagai,K. and Otsuki,T.  
TITLE Primer for synthesizing full-length cDNA and use thereof  
JOURNAL Patent: JP 2002191363-A 13123 09-JUL-2002;  
HELIX RESEARCH INSTITUTE  
COMMENT OS Homo sapiens (human)  
FN JP 2002191363-A/13123  
PD 09-JUL-2002  
PF 28-JUL-2000 JP 2000280990

PI TOSHIO OTA, TAKAO ISOGAI, TETSUO NISHIKAWA, KOJI HAYASHI, KAORU  
 PI SAITO,  
 PI JUNICHI YAMAMOTO, SHIZUKO ISHII, TOMOYASU SUGIYAMA, AI WAKAWATSU,  
 PI KEIICHI NAGAI, TETSUJI OTSUKI  
 PC  
 C12N15/09, C07K14/47, C07K16/18, C12N1/15, C12N1/19, C12N1/21, C12N5/ PC  
 10, C12P21/02, C12Q1/68//C12P21/08, C06F17/30, C12N15/00, C12N5/00 CC  
 Primer for synthesizing full-length cDNA and use thereof PH Key  
 Location/Qualifiers  
 FT CDS (42) . . (2453) .

## FEATURES

source  
 1. .2993  
 /organism="Homo sapiens"  
 /mol\_type="genomic DNA"  
 /db\_xref="taxon:9606"

## ORIGIN

## Alignment Scores:

Pred. No.: 2,55e-314 Length: 2993  
 Score: 295.00 Matches: 681  
 Percent Similarity: 98.84% Conservations: 0  
 Best Local Similarity: 98.84% Mismatches: 4  
 Query Match: 40.08% Indels: 8  
 DB: 6 Gaps: 0

US-10-045-815-4 (1-736) x BD158280 (1-2993)

Qy 1 MetAlaValArgAlaLeuLysLeuLeuThrThrLeuLeuAlaValAlaAlaSer 20  
 Db 42 ATGGCGGTACGCGGTGGAAGCTGCTGACCACTGCTGGCTGTGCGCGCTGCGCTCC 101  
 Qy 21 GlnAlaGluValGluSerGluAlaGlyTrpGlyMetValThrProAspLeuLeuPheAla 40  
 Db 102 CAAGCCGAGTCGAGTCCGAGGAGGATGGGCGATGGTACGCTGATCTGCTCTTGGCC 161  
 Qy 41 GluGlyThrAlaAlaTyAlaArgGlyAspTrpProGlyValValLeuSerMetGluArg 60  
 Db 162 GAGGGGACCGCAGCTACGCGCGGGGACCTGGGCGGGTGTCTCTGAGCATGGAACGG 221  
 Qy 61 AlaLeuArgSerArgAlaAlaLeuArgAlaLeuArgLeuArgCysArgThrGlnCysAla 80  
 Db 222 GCGTGTGCGCTCCGGGGAGCCCTCCGCGCCCTTCGCGCTGCGCTGCGCGCACCCAGTGTGCC 281  
 Qy 81 AlaAspPheProTrpGluLeuAspProAspTrpSerProSerProAlaGlnAlaSerGly 100  
 Db 282 GCCGACTTCCGCTGGAGCTGGACCCCGACTGGTCCCGCCAGCCCGCGCGCTCGGGC 341  
 Qy 101 AlaGly-AlaLeuArgAspLeuSerPhePheGlyGlyLeuLeuArgArgAlaAlaCysLe 120  
 Db 342 CGCCG- CGCCCTGCGCGACCTGAGCTTCTTCGGGGGCGCTTCTGCGCGCTGCGCTGCT 400  
 Qy 120 uArgArgCysLeuGlyProProAlaAlaHisSerLeuSerGluGluMetGluLeuGluPh 140  
 Db 401 GCGCGCTGCTCGCGCGCGCGCGCGCCACTCGCTCAGCGAGAGATGGAGTGGAGTT 460  
 Qy 140 eArgLysArgSerProTyAsnTyLeuGlnValAlaTyPheLysLeuLeuLysLeuG 160  
 Db 461 CGCGAAGGGAGCCCTCAACTACTCGAGTGGCTGCTTCAAGATCAACAAGTTGGA 520  
 Qy 160 uLysAlaValAlaAlaAlaHisThrPhePheValGlyAsnProGluHisMetGluMetG 180  
 Db 521 GAAAGCTGTGTGTCAGACACACACTTCTTGTGGGGGAACTCTGAGCACATGGAAATGCA 580  
 Qy 180 nGlnAsnLeuAspTyTrpGlnThrMetSerGlyValLysGluAlaAspPheLysAspLe 200  
 Db 581 GCAGAACCTAGACTATTACCAAAACCATGCTGGAGTGAAGGAGCGGACTTCAAGGATCT 640  
 Qy 200 uGluThrGlnProHisMetGlnGluPheArgLeuGlyValArgLeuTyTrpSerGluG 220  
 Db 641 TGAGACTCAACCCCATATGCAAGAAATTCGACTGGAGTGGAGTCTACTCAGAGGAACA 700  
 Qy 220 nProGlnGluAlaValProHisLeuGluAlaLeuGlnGluTyTrpPheValAlaTyG 240

Db 701 GCCACAGGAAGCTGTGCCACCACTAGCGCGCGCTGCAAGAACTATTGTGGCCTATGA 760  
 Qy 240 uGluCysArgAlaLeuCysGluGlyProTyAspTyTrpAspGlyTyTrpAsnTyLeuGluTy 260  
 Db 761 GGAGTGGCGTGGCTCTGCGAAGGGCCCTATGACTACGATGGCTCAACTACTCTGAGTA 820  
 Qy 260 rAsnAlaAspLeuPheGlnAlaAlaThrAspHisTyTrpGlnValLeuAsnCysLysG 280  
 Db 821 CAACGCTGACCTCTTCCAGGCCATCAGATCATTACATCCAGGTCCTCACTGTAAGCA 880  
 Qy 280 nAsnCysValThrGluLeuAlaSerHisProSerArgGluLysProPheGluAspPheLe 300  
 Db 881 GAACCTGTGTCAGGAGCTTGTCTCCACCCAAAGTCGAGAGAAGCCCTTTGAAGACTTCT 940  
 Qy 300 uProSerHisTyAsnTyTrpLeuGlnPheAlaTyTrpAsnLleGlyAsnTyTrpGlnAl 320  
 Db 941 CCCATCGCATTAATATCTGAGTTTGCCTTACTATAACATTTGGAAATATACACAGC 1000  
 Qy 320 aGly-GluCysAlaLysThrTyTrpLeuLeuPhePheProAsnAspGluValMetAsnGlnA 340  
 Db 1001 TGT-TGAATGTGCCAAGACTATCTTCTTCTTCCCAATGACGAGGTGATGAACCAA 1059  
 Qy 340 snLeuAlaTyTrpAlaAlaMetLeuGlyGluGluHisThrArgSerLleGlyProArgG 360  
 Db 1060 ATTGGCTTATTATGCAAGCTATGCTTGGAGAAGAACACACACAGATCCATCGGCCCGCTG 1119  
 Qy 360 lLysAlaLysGluTyArgGlnArgSerLeuLeuGluLysGluLeuLeuPhePheAlaT 380  
 Db 1120 AGAGTCCCAAGAGATACCGACGAGGCTTACTTGGAAAAGAACTGCTTTTCTTCTGCTT 1179  
 Qy 380 yTrpValPheGlyLleProPheValAspProAspSerTrpThrProGlu-GluValLle 399  
 Db 1180 ATGATGTTTTTGAATTTCCCTTTGTGATCCGGATTCATGGATCCAGG-AGAAGTGATT 1238  
 Qy 400 ProLysArgLeuGlnGlnLysGlnLysSerGluArgGluThrAlaValArgLleSerGln 419  
 Db 1239 CCACAGAGATTCCACAGAGAAACAGAGTCAGAACCGGAAACACCGCTACGCACTCCCCAG 1298  
 Qy 420 GluLleGlyAsnLeuMetLysGluLleGluThrLeuValGluGluLysThrLysGluSer 439  
 Db 1299 GAGATTGGGNACTTATGAGGNAATCGAGACCTTGTGGAGAGAGAACCAAGAGATCA 1358  
 Qy 440 LeuAspValSerArgLeuThrArgGluGlyGlyProLeuLeuTyrcLudlylleserLeu 459  
 Db 1359 CTGGATGTGAGCAGATGACCCGGGAAGGTGGCCCCCTCTGTATGAAGGCGATCAGTCTC 1418  
 Qy 460 ThrMetAsnSerLysLeuLeuAsnGlyTyTrp-GlnArgValValMetAspGlyValLleSe 479  
 Db 1419 ACCATGAACTCCAAACTCTGTATGGTTC-CCAGCGGGTGGTGGATGGAGCGGTAATCTC 1477  
 Qy 479 rAspHisGluCysGlnGluLeuGlnArgLeuThrAsnValAlaAlaThrSerGlyAspGl 499  
 Db 1478 TGACCACGAGTGTGAGGAGCTGCAGAGACTGACCAATGTGGCAGCAACCTCAGGAGATGG 1537  
 Qy 499 yTyArgGlyGlnThrSerProHisThrProAsnGluLysPheTyTrpGlyValThrValPh 519  
 Db 1538 CTACCGGGGTGAGACTCTCCCACTATCTCCCAATGAAAAGTCTATGGTGTACTGTCTT 1597  
 Qy 519 eLysAlaLeuLysLeuGlyGlnGluLysValProLeuGlnSerAlaHisLeuTyTrp 539  
 Db 1598 CAAGCCCTCAAGCTGGGCAAGAGGCAAGAGTTCTCTCTGCAGAGTGCACCTGTACTA 1657  
 Qy 539 rAsnValThrGluLysValArgArgLleMetGluSerTyPheArgLeuAspThrProLe 559  
 Db 1658 CAACGTGACGGGAAGGTGCGCGCATCATGAGTCTTACTTCCGCTCGATACGCCCT 1717  
 Qy 559 uTyTrpPheSerTySerHisLeuValCysArgThrAlaLleGluGluValGlnAlaGluAr 579  
 Db 1718 CTACTTTTCTACTCTCATCTGGTGTGCGGCACTGCGCATCGAAGAGTCCAGCAGAGAG 1777  
 Qy 579 gLysAspAspSerHisProValHisValAspAsnCysLleLeuAsnAlaGluThrLeuVa 599

```
Db 1778 GAAGGATGATGTCATCCAGTCCAGTGGACAACTGCATCCTGAATGCCGAGACCTCGT 1837
QY 599 ICysValIysGluProProAlaTyrThrPheArgAspTyrSerAlaIleLeuTyrLeuAs 619
Db 1838 GGTGTCTCAAAGAGCCCCCAGGCTACACCTCCGCGACTACAGCGCCATCCTTTACCTAAA 1897
QY 619 nGlyAspPheAspGlyGlyAsnPheTyrPheThrGluLeuAspAlaIysThrValThrAl 639
Db 1898 TGGGACTTCGATGCGGAAACTTTTATTCTCACTGAACCTGGATGCCAAGACCGTGACGGC 1957
QY 639 aGluValGlnProGlnCysGlyArgAlaValGlyPheSerSerGlyThrGluAsnProHi 659
Db 1958 AGAGGTGCAGCCTCAGTGTGAAGAGCGCGTGGATTCTCTCAGGCACCTGAAACCACCA 2017
QY 659 sGlyValIysAlaValThrArgGlyGlnArgCysAlaIleAlaLeuTrpPheThrLeuAs 679
Db 2018 TGGAGTGAAGGCTGTACACAGGGGGCAGCGCTGTGCCATCCGCTGTGGTTTCACTTGA 2077
QY 679 pProArgHisSerGluArg 685
Db 2078 CCTCGACACAGCGAGCGG 2096
```

Search completed: July 18, 2004, 21:38:07  
Job time : 10818 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model  
Run on: August 17, 2004, 10:09:10 ; Search time 56 Seconds  
(without alignments)  
3713.481 Million cell updates/sec

Title: US-10-045-815-4  
Perfect score: 736  
Sequence: 1 MAVRALKILTLAVAAAS.....PPEFAQESLGGSESKPKDEL 736

Scoring table: OLIGO  
Gapop 60.0 , Gapext 60.0

Searched: 1586107 seqs, 282547505 residues

Word size : 0

Total number of hits satisfying chosen parameters: 122046

Minimum DB seq length: 0  
Maximum DB seq length: 8

Post-processing: Listing first 45 summaries

Database : A\_Geneseq\_29Jan04.\*  
1: Geneseqp1980s.\*  
2: Geneseqp1990s.\*  
3: Geneseqp2000s.\*  
4: Geneseqp2001s.\*  
5: Geneseqp2002s.\*  
6: Geneseqp2003as.\*  
7: Geneseqp2003bs.\*  
8: Geneseqp2004s.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	6	0.8	7	4	AAB72810	Antibacte
2	6	0.8	7	4	AAB72814	Antibacte
3	6	0.8	7	4	AAB72822	Antibacte
4	6	0.8	8	2	AAR12310	ID2 plasm
5	6	0.8	8	2	AAR43245	Cytotoxic
6	6	0.8	8	2	AAR78840	P. falcip
7	6	0.8	8	2	AAR78840	Peptide f
8	6	0.8	8	2	AAY03662	Amino aci
9	6	0.8	8	2	AAB23655	Cytotoxic
10	6	0.8	8	7	ABE80013	Malaria C
11	5	0.7	5	1	AAE90014	Peptide f
12	5	0.7	5	2	AAW00426	Interleuk
13	5	0.7	5	3	AAV63463	Protocad
14	5	0.7	5	3	AAV6279	Linker us
15	5	0.7	5	3	AAE21929	Oligopept
16	5	0.7	5	5	AAE21931	Oligopept
17	5	0.7	5	5	AB46172	Desmoglei
18	5	0.7	6	1	AAE30276	Sequence
19	5	0.7	6	2	AAR11592	Fibrinoge
20	5	0.7	6	2	AAR10402	Fibrinoge
21	5	0.7	6	2	AAR57812	RGD contg
22	5	0.7	6	2	AAR57813	RGD contg
23	5	0.7	6	2	ADD94949	Platelet
24	5	0.7	6	2	ADD94957	Platelet
25	5	0.7	6	2	ADD94948	Platelet

26	S	0.7	6	2	ADD94950	Platelet
27	5	0.7	6	2	ADD94945	Platelet
28	5	0.7	6	2	ADD94955	Platelet
29	5	0.7	6	2	ADD94958	Platelet
30	5	0.7	6	2	ADD94956	Platelet
31	5	0.7	6	2	ADD94959	Platelet
32	5	0.7	6	2	ADD94951	Platelet
33	5	0.7	6	2	ADD94954	Platelet
34	5	0.7	6	2	ADD94953	Platelet
35	5	0.7	6	2	ADD94947	Platelet
36	5	0.7	6	2	ADD94946	Platelet
37	5	0.7	6	2	ADD94952	Platelet
38	5	0.7	6	2	AAR95089	Silk fibr
39	5	0.7	6	2	AAR32886	Peptide B
40	5	0.7	6	2	AAW26329	Silk fibr
41	5	0.7	6	2	AAW64612	Human poi
42	5	0.7	6	2	AAW53502	Peptide r
43	5	0.7	6	2	AAW57229	OA-519 pe
44	5	0.7	6	2	AAW23829	Human thr
45	5	0.7	6	2	AAW96753	Peptide f

ALIGNMENTS

RESULT 1  
AAB72810  
ID AAB72810 standard; peptide; 7 AA.  
XX  
XX AAB72810;  
XX  
XX 09-MAY-2001 (first entry)  
XX  
XX Antibacterial peptide SEQ ID NO: 31.  
DE  
XX Antibacterial compound; Gram-negative bacterium; pilus; chaperone;  
KW biofilm; disease treatment; bacterial infection.  
XX  
XX Synthetic.  
XX  
XX WO200110386-A2.  
XX  
XX 15-FEB-2001.  
XX  
XX 11-AUG-2000; 2000WO-US022087.  
XX  
XX 11-AUG-1999; 99US-0148280P.  
XX  
XX (UNIW ) UNIV WASHINGTON.  
XX  
XX Hultgren SJ, Sauer FG, Waksman G, Fuetterer X;  
XX  
XX WPI; 2001-226496/23.  
XX  
XX An isolated compound for inhibiting pilus assembly.  
XX  
XX Claim 12; Page 138; 144pp; English.

The present invention provides antibacterial compounds which are able to interfere with Gram-negative bacteria pilus formation and assembly, and pilus interaction with chaperone proteins. These are useful in the treatment of bacterial infection, and in the prevention of biofilm formation. They are particularly useful against *Escherichia coli*, *Haemophilus influenzae*, *Salmonella enteritidis*, *S. typhimurium*, *Bordetella pertussis*, *Yersinia enterocolitica*, *Helicobacter pylori* and *Klebsiella pneumoniae*

XX Sequence 7 AA;

Query Match 0.8%; Score 6; DB 4; Length 7;  
Best Local Similarity 100.0%; Pred. No. 1.4e+06;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 305 NYLQFA 310  
 DB 1 NYLQFA 6

RESULT 2  
 ID AAB72814 standard; peptide; 7 AA.  
 AC AAB72814;  
 XX 09-MAY-2001 (first entry)  
 DT Antibacterial peptide SEQ ID NO: 35.  
 DE Antibacterial compound; Gram-negative bacterium; pilus; chaperone;  
 KW biofilm; disease treatment; bacterial infection.  
 XX Synthetic.  
 OS WO200110386-A2.  
 PN 15-FEB-2001.  
 PD 11-AUG-2000; 2000WO-US022087.  
 XX 11-AUG-1999; 99US-0148280P.  
 XX (UNIW ) UNIV WASHINGTON.  
 PA Hultgren SJ, Sauer FG, Waksman G, Fuetterer K;  
 PI WPI; 2001-226496/23.  
 XX An isolated compound for inhibiting pilus assembly.  
 DR Claim 12; Page 141; 144pp; English.  
 XX The present invention provides antibacterial compounds which are able to  
 CC interfere with Gram-negative bacteria pilus formation and assembly, and  
 CC pilus interaction with chaperone proteins. These are useful in the  
 CC treatment of bacterial infection, and in the prevention of biofilm  
 CC formation. They are particularly useful against *Escherichia coli*,  
 CC *Haemophilus influenzae*, *Salmonella enteritidis*, *S. typhimurium*,  
 CC *Bordetella pertussis*, *Yersinia enterocolitica*, *Helicobacter pylori* and  
 CC *Klebsiella pneumoniae*  
 XX Sequence 7 AA;  
 SQ

Query Match 0.8%; Score 6; DB 4; Length 7;  
 Best Local Similarity 100.0%; Pred.No. 1.4e+06;  
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 305 NYLQFA 310  
 DB 1 NYLQFA 6

RESULT 4  
 AAB72814  
 ID AAB72814 standard; protein; 8 AA.  
 XX AAB72814;  
 AC AAB72814;  
 XX 25-MAR-2003 (revised)  
 DT 29-AUG-1991 (first entry)  
 XX ID2 plasmodium surface peptide.  
 DE Immunogenic determinant; circumsporozoite; CS; vaccine; malaria; hybrid.  
 KW Plasmodium falciparum.  
 OS EP432965-A.  
 PN 19-JUN-1991.  
 PD 06-DEC-1990; 90EP-00313257.  
 PF 08-DEC-1989; 89US-00447746.  
 XX (SMK ) SMITHKLINE BEECHAM.  
 PA (USSA ) US SEC OF ARMY.  
 PA (BIOM-) BIOMEDICAL RES INST.  
 PA (GROS/) GROSS M S.  
 XX Gross MS, Gordon DM, Hollingdal MR;  
 PI WPI; 1991-179771/25.  
 DR Polypeptide comprising immunogenic determinants from *P falciparum* - for  
 XX vaccine against malaria infection in humans.

QY 305 NYLQFA 310  
 DB 1 NYLQFA 6

RESULT 3  
 AAB72822  
 ID AAB72822 standard; peptide; 7 AA.  
 AC AAB72822;  
 XX 09-MAY-2001 (first entry)  
 DT Antibacterial peptide SEQ ID NO: 43.  
 DE Antibacterial compound; Gram-negative bacterium; pilus; chaperone;  
 KW biofilm; disease treatment; bacterial infection.  
 XX Synthetic.  
 OS

Query Match 0.8%; Score 6; DB 4; Length 7;  
 Best Local Similarity 100.0%; Pred.No. 1.4e+06;  
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 305 NYLQFA 310  
 DB 1 NYLQFA 6

RESULT 3  
 AAB72822  
 ID AAB72822 standard; peptide; 7 AA.  
 AC AAB72822;  
 XX 09-MAY-2001 (first entry)  
 DT Antibacterial peptide SEQ ID NO: 43.  
 DE Antibacterial compound; Gram-negative bacterium; pilus; chaperone;  
 KW biofilm; disease treatment; bacterial infection.  
 XX Synthetic.  
 OS

PS Claim 2; Page 16; 18pp; English.

XX The peptide is a sporozoite neutralising epitope from the 2nd flank- ing  
 CC region of the CS protein of plasmodium. It can be used in a vaccine for  
 CC protection against malaria. The peptide is pref. linked to a 2nd peptide  
 CC from the 1st flanking domain and the resulting polypeptide fused to a  
 CC carrier protein, e.g. tetanus toxoid, dip- theria toxin or cholera B  
 CC toxin. A preferred vaccine comprises 81 N-terminal AAs of the influenza  
 CC virus nonstructural protein 1 (NS181), fused, via a synthetic linker, to  
 CC a 1st flanking region of the CS protein, which is itself fused to a 2nd  
 CC flanking region of the CS protein. See also AAR12306-R12311 and AAR13175-  
 CC R13179. (Updated on 25-MAR-2003 to correct PA field.)

XX Sequence 8 AA;  
 SQ Query Match 0.8%; Score 6; DB 2; Length 8;  
 Best Local Similarity 100.0%; Pred. No. 1.4e+06;  
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 731 KPDEL 736  
 Db |||||  
 2 KPDEL 7

RESULT 5  
 AAR43245  
 ID AAR43245 standard; peptide; 8 AA.  
 AC AAR43245;  
 XX 25-MAR-2003 (revised)  
 DT 04-MAY-1994 (first entry)  
 XX Cytotoxic T lymphocyte recognition/induction peptide.  
 DE CTL; vaccine; malaria; specific antigen-derived.  
 XX Synthetic.  
 OS WO9320103-A2.  
 XX 14-OCT-1993.  
 PD 05-APR-1993; 93WO-GB000711.  
 XX 03-APR-1992; 92GB-00008068.  
 PR 20-AUG-1992; 92GB-00017704.  
 XX (ISIS-) ISIS INNOVATION LTD.  
 PA Hill AV, Gotch FM, McMichael AJ, Whittle HC;  
 PI WPI; 1993-336833/42.  
 DR Peptide(s) recognising or inducing cytotoxic T lymphocytes - useful in  
 XX vaccines against malaria or HIV-2, derived from specific antigen and  
 PT human leukocyte antigen contg. class I restricted epitope.  
 XX Claim 6; Page 30; 35pp; English.  
 PS The sequence is that of peptide cp26 which is recognised by, or can  
 XX induce, cytotoxic T lymphocytes. It may be useful in vaccines against  
 CC malaria. (Updated on 25-MAR-2003 to correct PN field.)

XX Sequence 8 AA;  
 SQ Query Match 0.8%; Score 6; DB 2; Length 8;  
 Best Local Similarity 100.0%; Pred. No. 1.4e+06;  
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 731 KPDEL 736  
 Db |||||  
 1 KPDEL 6

RESULT 6  
 AAR78840  
 ID AAR78840 standard; peptide; 8 AA.  
 XX AAR78840;  
 AC AAR78840;  
 XX 25-MAR-2003 (revised)  
 DT 27-MAR-1996 (first entry)  
 XX P. falciparum circumsp. 368-375 cytotoxic T lymphocyte epitope.  
 DE Circump. 368-375; cytotoxic T; CTL; epitope; helper T; HTL; cell;  
 KW lymphocyte; viruses; parasites; tumours; antigens; treatment;  
 KW disease prevention.  
 XX Plasmodium falciparum.  
 OS WO9522317-A1.  
 XX 24-AUG-1995.  
 PD 16-FEB-1995; 95WO-US002121.  
 XX 16-FEB-1994; 94US-00197484.  
 PR (CYTE-) CYTEL CORP.  
 XX Vitello MA, Chesnut RW, Sette AD, Celis E, Grey H;  
 PI WPI; 1995-302545/39.  
 DR Compsn. inducing cytotoxic T lymphocyte response to pref. viral,  
 XX bacterial, parasitic or tumour antigens - useful in the treatment and  
 PT prevention of diseases associated with the antigen e.g. hepatitis B.  
 XX Disclosure; Page 17; 109pp; English.  
 CC A compsn. which induces a cytotoxic T lymphocyte (CTL) response to an  
 CC antigen (Ag) in a mammal comprises, a CTL Ag response inducing peptide  
 CC (i.e. AAR78824-R78853) and a lipid conjugated helper T cell inducing  
 CC peptide. The compsn. induces a CTL response to bacterial, viral or tumour  
 CC Ags, and is therefore useful in the treatment and prevention of diseases  
 CC associated with the Ag. (Updated on 25-MAR-2003 to correct PI field.)

XX Sequence 8 AA;  
 SQ Query Match 0.8%; Score 6; DB 2; Length 8;  
 Best Local Similarity 100.0%; Pred. No. 1.4e+06;  
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 731 KPDEL 736  
 Db |||||  
 1 KPDEL 6

RESULT 7  
 AAW54663  
 ID AAW54663 standard; peptide; 8 AA.  
 XX AAW54663;  
 AC AAW54663;  
 XX 25-SEP-1998 (first entry)  
 DT Peptide from P. falciparum CSP 368-375.  
 XX Mannose; antigen; antigen-presenting cell; mannosylated peptide; T cell;  
 KW vaccine; treatment.  
 XX Synthetic.  
 OS WO9813378-A1.  
 XX PN



XX PD 02-APR-1998.  
 XX XX 25-SEP-1997; 97WO-NL000536.  
 XX PF  
 XX PR 26-SEP-1996; 96EP-00202701.  
 XX XX (UYLE-) RIJKSUNIV LEIDEN.  
 XX PA Koning F, Drijfhout JW;  
 XX PI WPI; 1998-230631/20.  
 XX DR  
 XX PT Increasing uptake and presentation of antigen(s) - by adding mannose  
 PT residue(s) to antigen for increasing T cell response, useful in, e.g.  
 PT vaccines against viral infection(s).  
 XX XX Disclosure: Page 29; 47pp; English.  
 XX CC The peptides AAW54559-W54809 are examples of peptides to which at least 1  
 CC (preferably 2) mannose can be attached to increase their uptake as  
 CC antigens by antigen-presenting cells. Uptake of agonist mannoseylated  
 CC peptides will increase the T cell response, whereas uptake of antagonist  
 CC peptides blocks the T cell response. Blocking binding of immunogenic  
 CC autoantigens can be used in treatment of type I diabetes, rheumatoid  
 CC arthritis, graft rejection etc., also to induce T-cell non-  
 CC responsiveness. Vaccines containing mannoseylated antigen are used to  
 CC prevent or treat infections by, e.g. bacteria, viruses, fungi, helminths  
 CC and parasites  
 XX CC  
 XX SQ Sequence 8 AA;  
 Query Match 0.8%; Score 6; DB 2; Length 8;  
 Best Local Similarity 100.0%; Pred. No. 1.4e+06;  
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 731 KPKDEL 736  
 Db |||||  
 1 KPKDEL 6  
 RESULT 8  
 AAY03662  
 ID AAY03662 standard; peptide; 8 AA.  
 XX AC AAY03662;  
 XX DT 07-JUN-1999 (first entry)  
 XX DE Amino acid sequence of the malaria (M) string CTL epitope Cp26.  
 XX CD8+ T-cell; immune response; antigen; priming composition; CTL; epitope;  
 KW cytotoxic T lymphocyte; boosting; poxvirus vector; PVV; pathogen; tumour;  
 KW malaria; parasite; P. falciparum; viral; bacterial; parasitic; cancer;  
 KW melanoma; HIV; breast; colon; vaccination.  
 XX OS Plasmodium falciparum.  
 XX XX WO9856919-A2.  
 XX PD 17-DEC-1998.  
 XX PF 09-JUN-1998; 98WO-GB001681.  
 XX PR 09-JUN-1997; 97GB-00011957.  
 XX XX (ISIS-) ISIS INNOVATION LTD.  
 XX PA McMichael AJ, Hill AVS, Gilbert SC, Schneider J, Plebanski M;  
 XX PI Hanke T, Smith GL, Blanchard T;  
 XX DR WPI; 1999-070325/06.  
 DR N-PSDB; AAX29202.

XX PT Generating CD8-positive T cell response to target antigen using  
 PT recombinant poxvirus - for treating or preventing malaria and HIV  
 PT infection, also epitope strings from Plasmodium and HIV.  
 XX PS Claim 38; Page 18; 85pp; English.  
 XX CC The invention relates to methods and reagents for generating a protective  
 CC CD8+ T-cell immune response against at least one target antigen. The kits  
 CC of the invention comprises (i) as priming composition, a source of one or  
 CC more CD8+ T-cell (cytotoxic T lymphocytes-(CTL)) epitopes of the target  
 CC antigen, plus a carrier and (ii) as boosting composition a source of CTL  
 CC epitopes, with at least one CTL epitope the same as used in (i), with  
 CC this source being a non-replicating or replication-impaired recombinant  
 CC poxvirus vector (PVV) plus a carrier. If the source of CTL epitopes in  
 CC (i) is a viral vector, then the vector in (ii) is from a different virus.  
 CC The kits are used to generate an immune response (prophylactic or  
 CC therapeutic) against pathogens or tumours, specifically against malaria  
 CC parasites such as P. falciparum, or HIV, and also many other bacterial,  
 CC viral or parasitic pathogens. The kits are also used for protective  
 CC response against melanoma and cancer of breast or colon, and generally  
 CC wherever a strong CD8+ response is protective. The boosting composition  
 CC may be used alone to boost a naturally primed response against malaria.  
 CC The specified PVV provide an excellent booster effect, better than that  
 CC from wild-type poxvirus, resulting in complete rather than partial  
 CC protection against sporozoite challenge. Also PVV are safer to use than  
 CC wild-type virus. Sequences AAY03661-680 represent CTL peptide epitopes of  
 CC the malaria (M) string  
 XX CC  
 XX SQ Sequence 8 AA;  
 Query Match 0.8%; Score 6; DB 2; Length 8;  
 Best Local Similarity 100.0%; Pred. No. 1.4e+06;  
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 731 KPKDEL 736  
 Db |||||  
 1 KPKDEL 6  
 RESULT 9  
 AAB23655  
 ID AAB23655 standard; peptide; 8 AA.  
 XX AC AAB23655;  
 XX DT 05-JAN-2001 (first entry)  
 XX DE Cytotoxic T lymphocyte (CTL) epitope SEQ ID NO:7.  
 XX KW ATPase; Hsp70; heat shock protein; cytotoxic T lymphocyte; CTL;  
 KW immune response; infectious disease; malaria; cytotoxic T cell;  
 KW cytostatic; immunostimulant; cellular immune response inducer;  
 KW protozoacide; leukaemia; cancer.  
 XX OS Homo sapiens.  
 XX XX WO200049041-A1.  
 XX PD 24-AUG-2000.  
 XX PF 18-FEB-2000; 2000WO-JP000941.  
 XX PR 19-FEB-1999; 99TP-00041535.  
 XX XX (SUME) SUMITOMO ELECTRIC IND CO.  
 XX PA Shinbara N, Udono H, Yui K;  
 XX PI WPI; 2000-543748/49.  
 XX DR  
 XX PT Fused protein capable of inducing cellular immune response, useful as  
 PT active ingredient for drug compositions in preventing and/or treating

PT infectious diseases such as malaria or cancer.

PS Claim 5; Page 52; 72pp; Japanese.

XX The present invention describes a fused protein (I) prepared from a peptide containing a CTL (cytotoxic T lymphocyte) epitope recognised by cytotoxic T cells and a protein containing the ATPase domain of a heat shock protein. Also described are: (1) a drug composition containing (I) as active ingredient; (2) a DNA encoding (I); (3) an expression vector containing the DNA of (2); and (4) a transformant which can retain the expression vector of (3). (I) has cytostatic, immunostimulant and protozoacide activities, and can be used as a cellular immune response inducer. The protein is useful as an active ingredient for drug compositions in preventing and/or treating infectious diseases such as malaria or cancer e.g. to provide systemic immunity against leukaemia. The present sequence represents a specifically claimed CTL epitope for use in a fused protein of the present invention

XX SQ Sequence 8 AA;

Query Match 0.8%; Score 6; DB 3; Length 8;  
Best Local Similarity 100.0%; Pred. No. 1.4e+06;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 731 KPKDEL 736  
Db 1 KPKDEL 6

RESULT 10

AD890013  
ID ADE80013 standard; peptide; 8 AA.

AC ADE80013;

DT 29-JAN-2004 (first entry)

XX Malaria CTL epitope in method to generate CD8+ T-cell immune response.

XX antimalarial; cytostatic; vaccine; immune response;  
KW non-hepadnaviral antigen; hepatitis B core particle; CD8+ T-cell;  
KW epitope; poxvirus vector; cancer; malaria; epitope.  
XX Plasmodium falciparum.

OS WO2003066833-A2.

PN 14-AUG-2003.

PF 07-FEB-2003; 2003WO-US003897.

XX 08-FEB-2002; 2002US-0354963P.

XX (UUNY-) UNIV NEW YORK MEDICAL CENT.

XX Zavala F, Birkett AJ;

XX WPI; 2003-748124/70.

XX Generating an immune response against a non-hepadnaviral antigen in a mammal, useful for treating or preventing cancer or malaria, by administering a priming component comprising a recombinant hepatitis B core particle.

PS Disclosure; SEQ ID NO 69; 85pp; English.

XX The invention relates to a method of generating an immune response against a non-hepadnaviral antigen in a mammal by administering (to the mammal) at least 1 dose of a priming component comprising a recombinant hepatitis B core particle (RHEP) (which is a carrier for 1 or more non-hepadnaviral CD8+ T-cell epitopes of the antigen). The method may be supplemented by the use of a boosting stage comprising a non-replicating or replication-impaired recombinant poxvirus vector. The method is useful

CC for generating an immune response against a non-hepadnaviral antigen in a mammal for treating or preventing cancer or malaria. This sequence represents a Plasmodium falciparum CTL peptide used to generate an immune response against a Plasmodium peptide.

XX SQ Sequence 8 AA;

Query Match 0.8%; Score 6; DB 7; Length 8;  
Best Local Similarity 100.0%; Pred. No. 1.4e+06;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 731 KPKDEL 736  
Db 1 KPKDEL 6

RESULT 11

AD890014  
ID AAP90014 standard; protein; 5 AA.

XX AAP90014;

AC 25-MAR-2003 (revised)

DT 01-NOV-1989 (first entry)

XX Peptide from HLA Class II beta region contg. residue corresp. tp DQ-beta protein.

XX Peptide; DQ-beta locus of HLA class II beta genes; allele-specific.

OS Homo sapiens.

PN WO8904875-A.

XX 01-JUN-1989.

XX 14-NOV-1989; 89WO-US004067.

XX 17-NOV-1987; 87US-00121519.

XX (CETU) CETUS CORP.

XX Erlich HA, Horn GT;

XX WPI; 1989-178393/24.

XX Marker DNA sequences from HLA class-II beta region - detect amino acid 57 codon of dq-beta protein to detect auto-immune susceptibility.

PS Claim 14; Page 59; 72pp; English.

XX Peptide contg. an epitope which has an amino acid residue corresp. to position 57 of a DQ-beta protein from the HLA class II beta genes, used to raise antibodies to (in)directly detect the identity of codon-57 of the DQ-beta protein sequence. Pref. codon-57 is selected from Ala, Val and Asp. Used to detect autoimmune diseases, esp. diabetes mellitus, and Pemphigus vulgaris. (Updated on 25-MAR-2003 to correct PF field.)

XX SQ Sequence 5 AA;

Query Match 0.7%; Score 5; DB 1; Length 5;  
Best Local Similarity 100.0%; Pred. No. 1.4e+06;  
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 125 GPPAA 129  
Db 1 GPPAA 5

RESULT 12

AAW00426  
ID AAW00426 standard; peptide; 5 AA.

XX

AC AAW00426;  
 XX 29-AUG-1996 (first entry)  
 DT  
 XX Interleukin-6 antagonist peptide.  
 DE  
 XX IL-6; antagonist; autoimmune disease.  
 KW  
 XX Synthetic.  
 OS  
 XX JF07324097-A.  
 XX 12-DEC-1995.  
 XX  
 XX 30-MAY-1994; 94JP-00117259.  
 XX  
 XX 30-MAY-1994; 94JP-00117259.  
 XX  
 XX (DAIL) DAICEL CHEM IND LTD.  
 PA (FUJI) FUJISAWA PHARM CO LTD.  
 PA  
 XX WPI; 1996-065476/07.  
 XX  
 XX Interleukin 6 antagonist - useful for treating autoimmune diseases.  
 PT  
 XX Claim 3, 6; Page 2, 3; 19pp; Japanese.  
 PS  
 XX New IL-6 antagonists are provided which are of formula X-W-Y, in which X  
 CC is H or an amino-protecting group; Y is OH or a carboxy-protecting  
 CC group, and W is a peptide containing all or part of the sequence as given  
 CC in AAW00401, AAW00402, AAW00403 or AAW00404, where any free mercapto  
 CC groups in the sequence are optionally protected. The present sequence is  
 CC a specifically preferred partial sequence of AAW00402 and is itself  
 CC claimed as a new chemical entity. The IL-6 antagonists are useful for  
 CC treating autoimmune diseases  
 XX  
 XX Sequence 5 AA;  
 SQ  
 Query Match 0.7%; Score 5; DB 2; Length 5;  
 Best Local Similarity 100.0%; Pred. No. 1.4e+06;  
 Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 68 LRALR 72  
 Db |||||  
 1 LRALR 5  
 RESULT 13  
 AAV63463  
 ID AAV63463 standard; peptide; 5 AA.  
 XX  
 AC AAV63463;  
 XX  
 DT 02-MAR-2000 (first entry)  
 DE  
 XX Protocadherin cell adhesion recognition cyclic peptide SEQ ID NO:2953.  
 XX  
 KW Modulation; nonclassical cadherin mediated cell adhesion; CAR;  
 KW inhibition; cadherin extracellular domain; cell adhesion recognition;  
 KW OB-cadherin; cadherin-5; cadherin-6; cadherin-7; cadherin-8; cadherin-12;  
 KW cadherin-14; cadherin-15; T-cadherin; PB-cadherin;  
 KW cadherin related neuronal receptor; LI-cadherin; protocadherin;  
 KW desmoglein; desmocollin; calcium binding; cancer; tumour; obesity;  
 KW rheumatoid arthritis; multiple sclerosis; diabetes; metastasis;  
 KW neurological disease; cyclic.  
 XX  
 OS Synthetic.  
 OS Homo sapiens.  
 XX  
 XX Key Location/Qualifiers  
 XX Modified-site 1.5  
 FT /note= "the terminal residues are condensed with each  
 FT other to form a cyclic peptide"  
 FT

XX WO9957149-A2.  
 XX 11-NOV-1999.  
 XX  
 XX 05-MAY-1999; 99WO-CA000363.  
 XX  
 XX 05-MAY-1998; 98US-00073040.  
 XX 06-NOV-1998; 98US-00187859.  
 XX 20-JAN-1999; 99US-00234395.  
 XX 08-MAR-1999; 99US-00264516.  
 XX  
 XX (ADHE-) ADHEREX TECHNOLOGIES INC.  
 XX  
 XX Blaschuk OW, Cour BJ, Byers S;  
 XX WPI; 2000-038791/03.  
 XX  
 XX New cadherin modulating agents, used for modulating nonclassical cadherin  
 CC -mediated functions for treating e.g. cancers, obesity, rheumatoid  
 CC arthritis, multiple sclerosis, diabetes or a neurological disease.  
 XX  
 XX Claim 84; Page 204; 252pp; English.  
 PS  
 XX The present invention describes cadherin modulating agents (MA)  
 CC comprising peptides which comprise a nonclassical cadherin cell adhesion  
 CC recognition (CAR) sequence. The MAs can be used for modulating  
 CC nonclassical cadherin-mediated functions. They can be used for e.g.  
 CC inhibiting adhesion of nonclassical-cadherin expressing cells in a  
 CC mammal, enhancing delivery of a drug to a tumour in a mammal, treating cancer in  
 CC a mammal, inhibiting metastasis of a cancer in a mammal, inhibiting  
 CC angiogenesis in a mammal, inducing apoptosis in a nonclassical cadherin-  
 CC expressing cell, preventing or treating obesity in a mammal, stimulating  
 CC blood vessel regression in a mammal, enhancing drug delivery to the  
 CC central nervous system, treating a demyelinating neurological disease,  
 CC increasing vasopermeability in a mammal, enhancing adhesion of  
 CC nonclassical cadherin-expressing cells, inhibiting synaptic stability in  
 CC a mammal, or preventing pregnancy in a mammal. They can also be used for  
 CC e.g. enhancing or directing neurite outgrowth, facilitating wound healing  
 CC or reducing scar tissue, or enhancing adhesion of foreign tissue in a  
 CC mammal. They can also be used for treating e.g. psoriasis, arthritis, age  
 CC related macular degeneration, multiple sclerosis and diabetes. The  
 CC products can also be used for detection and diagnosis and in bioassays.  
 CC AAY60592 to AAY64572 represent specifically claimed peptides, and  
 CC AAY64573 to AAY64583 and AAZ33183 to AAZ33186 represent sequences used in  
 CC the exemplification of the present invention  
 XX  
 XX Sequence 5 AA;  
 SQ  
 Query Match 0.7%; Score 5; DB 3; Length 5;  
 Best Local Similarity 100.0%; Pred. No. 1.4e+06;  
 Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 691 DDLVK 695  
 Db |||||  
 1 DDLVK 5  
 RESULT 14  
 AAY96279  
 ID AAY96279 standard; peptide; 5 AA.  
 XX  
 AC AAY96279;  
 XX  
 DT 11-AUG-2000 (first entry)  
 DE  
 XX Linker used to find IGF-1 receptor interacting proteins IIP cDNAs.  
 XX  
 XX IGF-1; IGF-1 receptor interacting protein; IIP; cancer; gene therapy;  
 KW tumour proliferation; signal transduction; protein-protein interaction;  
 KW diabetes; neurodegenerative disease; bone disease; agonist;  
 KW prostate cancer; breast cancer; linker.

```

XX OS Mammalia.
XX PN EPI006184-A1.
XX PD 07-JUN-2000.
XX PF 03-DEC-1998; 98EP-00122992.
XX PR 03-DEC-1998; 98EP-00122992.
XX PA (HOFF ) HOFFMANN LA ROCHE & CO AG F.
XX WPI; 2000-367969/32.
XX N-PSDB; AAA27351.
XX PT New IGF-1R-interacting protein for diagnosis of tumor progression and
XX treatment of tumors.
XX PS Disclosure; Fig 1; 31pp; English.
XX CC The present sequence is a peptide linker inserted between the LexA-
XX binding domain and IGF-1 receptor domain in a two-hybrid bait constructed
XX to screen a cDNA library for the mammalian IGF-1 receptor interacting
XX protein (IIP) genes. The proteins interact with IGF-1 receptor in the
XX cell, which is in turn overexpressed in many tumours. The use of the
XX proteins and genes in gene therapy and to identify agonists, antagonists
XX and antibodies will enable the diagnosis and treatment of disorders such
XX as diabetes, neurodegenerative disorders, bone diseases and cancers, in
XX particular prostate and breast cancers, as well as enabling the
XX inhibition of tumour progression. The genes are also useful for creating
XX animal models for these diseases
XX SQ Sequence 5 AA;
XX Query Match 0.7%; Score 5; DB 3; Length 5;
XX Best Local Similarity 100.0%; Pred. No. 1.4e+06;
XX Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
XX QY 139 EPRKR 143
XX |||||
XX Db 1 EPRKR 5
XX
XX RESULT 15
XX ID AAE21929
XX AC AAE21929 standard; peptide; 5 AA.
XX AC AAE21929;
XX DT 16-JUL-2002 (first entry)
XX DE Oligopeptide #5 used to generate plasmin-cleavable prodrug.
XX KW plasmin; prodrug; cancer; neoplastic disease; inflammatory disease;
XX tumour; infectious disease; chemosensitiser; dye; cytostatic.
XX OS Unidentified.
XX FH Key Location/Qualifiers
XX FT Misc-difference 1 /note= "D-form residue"
XX FT FT Misc-difference 5 /note= "Optionally linked to daunorubicin (DNR) or
XX FT doxorubicin (DOX) drug"
XX PN WO200215700-A1.
XX PD 28-FEB-2002.
XX PF 23-AUG-2001; 2001WO-US026476.
XX PR 24-AUG-2000; 2000US-0227686P.

```

```

XX (COUS ) COULTER PHARM INC.
XX Trouet A, Dubois V, Passioukov A;
XX WPI; 2002-339507/37.
XX PT New plasmin-cleavable prodrugs comprise an oligopeptide linked to a
XX stabilizing group and a therapeutic agent.
XX PS Example 1; Page 43; 67pp; English.
XX CC The present invention relates to plasmin-cleavable prodrugs comprising an
XX oligopeptide directly linked at a first attachment site to a stabilising
XX group that hinders cleavage of the oligopeptide by enzymes present in
XX whole blood and linked at a second attachment site, either directly or
XX through a linker, to a therapeutic agent capable of entering a target
XX cell. The prodrugs of the invention are useful for delivering therapeutic
XX agents, preferably alkylating agents, anthracyclines, antiproliferative
XX agents, camptothecins, chemotherapeutic agents, cyclosporins, enediyne,
XX dolastatins, epipodophyllotoxins, maytansinoids, naphthalimides, platinum
XX complexes, pteridines, rhodamines, sulfoximes, topoisomerase inhibitors,
XX taxanes, taxoids, tubulin binding agents or vinca alkaloids, especially
XX daunorubicin or doxorubicin. They are used to target cells in high-
XX plasmin environments so that the therapeutic agent is cleaved from the
XX prodrug in a form capable of entering the target cells, especially in the
XX treatment of cancer, neoplastic diseases, inflammatory diseases, tumours
XX and infectious diseases. Sequences of the invention can be used with
XX agents that act as chemosensitisers and dyes. The present sequence is an
XX oligopeptide used to generate plasmin-cleavable prodrugs of the invention
XX SQ Sequence 5 AA;
XX Query Match 0.7%; Score 5; DB 5; Length 5;
XX Best Local Similarity 100.0%; Pred. No. 1.4e+06;
XX Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
XX QY 5 ALKLL 9
XX |||||
XX Db 1 ALKLL 5
XX
XX Search completed: August 17, 2004, 10:15:20
XX Job time : 58 secs

```

This Page Blank (uspto)



US-08-318-856A-3

Query Match 0.8%; Score 6; DB 2; Length 8;  
 Best Local Similarity 100.0%; Pred.No. 3e+05;  
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 731 KPDEL 736

|||||  
 Db 1 KPDEL 6

RESULT 2

US-08-197-484-17  
 ; Sequence 17, Application US/08197484  
 ; Patent No. 6419931  
 ; GENERAL INFORMATION:  
 ; APPLICANT: VITIELLO, Maria A.  
 ; APPLICANT: CHESTNUT, Robert W.  
 ; APPLICANT: SETTE, Alessandro D.  
 ; APPLICANT: CELIS, Esteban  
 ; APPLICANT: GRAY, Howard  
 ; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR ELICITING  
 ; TITLE OF INVENTION: CTL IMMUNITY  
 ; NUMBER OF SEQUENCES: 13  
 ; CORRESPONDENCE ADDRESS:  
 ; ADDRESSEE: Townsend and Townsend Kourie and Crew  
 ; STREET: Steuart Street Tower, One Market Plaza  
 ; CITY: San Francisco  
 ; STATE: California  
 ; COUNTRY: US  
 ; ZIP: 94105-1493

COMPUTER READABLE FORM:  
 MEDIUM TYPE: Floppy disk  
 COMPUTER: IBM PC compatible  
 OPERATING SYSTEM: PC-DOS/MS-DOS  
 SOFTWARE: PatentIn Release #1.0, Version #1.25  
 CURRENT APPLICATION DATA:  
 APPLICATION NUMBER: US/08/197,484  
 FILING DATE: 16-FEB-1994  
 CLASSIFICATION: 424

PRIOR APPLICATION DATA:  
 APPLICATION NUMBER: US 07/935,811

FILING DATE: 26-AUG-1992

PRIOR APPLICATION DATA:  
 APPLICATION NUMBER: US 07/874,491

FILING DATE: 27-APR-1992

PRIOR APPLICATION DATA:  
 APPLICATION NUMBER: US 07/749,568

FILING DATE: 28-AUG-1991

ATTORNEY/AGENT INFORMATION:

NAME: Parmelee, Steven W.

REGISTRATION NUMBER: 31,990

REFERENCE/DOCKET NUMBER: 14137-26-4

TELEPHONE: (206) 467-9600

TELEFAX: (206) 623-6793

INFORMATION FOR SEQ ID NO: 17:

SEQUENCE CHARACTERISTICS:

LENGTH: 8 amino acids

TYPE: amino acid

STRANDEDNESS: unknown

TOPOLOGY: unknown

MOLECULE TYPE: peptide

US-08-197-484-17

Query Match 0.8%; Score 6; DB 4; Length 8;  
 Best Local Similarity 100.0%; Pred.No. 3e+05;  
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 731 KPDEL 736

|||||  
 Db 1 KPDEL 6

RESULT 3

US-09-454-204A-4  
 ; Sequence 4, Application US/09454204A  
 ; Patent No. 6663871  
 ; GENERAL INFORMATION:  
 ; APPLICANT: McMichael, Andrew  
 ; APPLICANT: Hill, Adrian V.S.  
 ; APPLICANT: Gilbert, Sarah C.  
 ; APPLICANT: Schneider, Jorg  
 ; APPLICANT: Plebanski, Magdalena  
 ; APPLICANT: Hanke, Tomas  
 ; APPLICANT: Smith, Geoffrey L.  
 ; APPLICANT: Blanchard, Tom  
 ; TITLE OF INVENTION: Methods and Reagents for Vaccination  
 ; TITLE OF INVENTION: Which Generate A CD8 T Cell Immune Response  
 ; FILE REFERENCE: 2907.1000-000  
 ; CURRENT APPLICATION NUMBER: US/09/454,204A  
 ; CURRENT FILING DATE: 1999-12-09  
 ; PRIOR APPLICATION NUMBER: PCT/GB98/01681  
 ; PRIOR FILING DATE: 1998-06-09  
 ; PRIOR APPLICATION NUMBER: GB 97 11957.2  
 ; PRIOR FILING DATE: 1997-06-09  
 ; NUMBER OF SEQ ID NOS: 78  
 ; SOFTWARE: FastSeq for Windows Version 4.0  
 ; SEQ ID NO 4  
 ; LENGTH: 8  
 ; TYPE: PRT  
 ; ORGANISM: Unknown  
 ; FEATURE:  
 ; OTHER INFORMATION: CTL Epitope of the Malaria String  
 ; US-09-454-204A-4

Query Match 0.8%; Score 6; DB 4; Length 8;  
 Best Local Similarity 100.0%; Pred.No. 3e+05;  
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 731 KPDEL 736

|||||  
 Db 1 KPDEL 6

RESULT 4

PCT-US95-02121-17  
 ; Sequence 17, Application PC/TUS9502121  
 ; GENERAL INFORMATION:  
 ; APPLICANT:  
 ; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR ELICITING  
 ; TITLE OF INVENTION: CTL IMMUNITY  
 ; NUMBER OF SEQUENCES: 153  
 ; COMPUTER READABLE FORM:  
 MEDIUM TYPE: Floppy disk  
 COMPUTER: IBM PC compatible  
 OPERATING SYSTEM: PC-DOS/MS-DOS  
 SOFTWARE: PatentIn Release #1.0, Version #1.25  
 CURRENT APPLICATION DATA:  
 APPLICATION NUMBER: PCT/US95/02121  
 FILING DATE: 16-FEB-1995  
 CLASSIFICATION:  
 PRIOR APPLICATION DATA:  
 APPLICATION NUMBER: US 08/197,484  
 FILING DATE: 16-FEB-1994  
 PRIOR APPLICATION DATA:  
 APPLICATION NUMBER: US 07/935,811  
 FILING DATE: 26-AUG-1992  
 PRIOR APPLICATION DATA:  
 APPLICATION NUMBER: US 07/874,491  
 FILING DATE: 27-APR-1992  
 PRIOR APPLICATION DATA:  
 APPLICATION NUMBER: US 07/827,682



; FILING DATE: 29-JAN-1992  
 ; PRIOR APPLICATION DATA:  
 ; APPLICATION NUMBER: US 07/749,568  
 ; FILING DATE: 26-AUG-1991  
 ; ATTORNEY/AGENT INFORMATION:  
 ; NAME: Parmelee, Steven W.  
 ; REGISTRATION NUMBER: 31,990  
 ; REFERENCE/DOCKET NUMBER: 14137-26-4PC  
 ; TELECOMMUNICATION INFORMATION:  
 ; TELEPHONE: (206) 467-9600  
 ; TELEFAX: (415) 543-5043  
 ; INFORMATION FOR SEQ ID NO: 17:  
 ; SEQUENCE CHARACTERISTICS:  
 ; LENGTH: 8 amino acids  
 ; TYPE: amino acid  
 ; STRANDEDNESS: unknown  
 ; TOPOLOGY: unknown  
 ; MOLECULE TYPE: peptide  
 ; PCT-US95-02121-17

Query Match 0.8%; Score 6; DB 5; Length 8;  
 Best Local Similarity 100.0%; Pred. No. 3e+05;  
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 731 KPKDEL 736  
 Db 1 KPKDEL 6

RESULT 5  
 US-09-187-859-2953  
 ; Sequence 2953, Application US/09187859A  
 ; Patent No. 6358920  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Blaschuk, Orest W.  
 ; APPLICANT: Gour, Barbara J.  
 ; TITLE OF INVENTION: COMPOUNDS AND METHODS FOR MODULATING NONCLASSICAL  
 ; TITLE OF INVENTION: CADHERIN-MEDIATED FUNCTIONS  
 ; FILE REFERENCE: 100086.407C1  
 ; CURRENT APPLICATION NUMBER: US/09/187,859A  
 ; CURRENT FILING DATE: 1998-11-06  
 ; NUMBER OF SEQ ID NOS: 4052  
 ; SOFTWARE: PatentIn Ver. 2.0  
 ; SEQ ID NO 2953  
 ; LENGTH: 5  
 ; TYPE: PRT  
 ; ORGANISM: Artificial Sequence  
 ; FEATURE:  
 ; OTHER INFORMATION: Representative cyclic modulating agent based on  
 ; OTHER INFORMATION: Protocadherin cell adhesion recognition sequence  
 US-09-187-859-2953

Query Match 0.7%; Score 5; DB 4; Length 5;  
 Best Local Similarity 100.0%; Pred. No. 3e+05;  
 Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 691 DDLVK 695  
 Db 1 DDLVK 5

RESULT 6  
 US-09-839-542B-2953  
 ; Sequence 2953, Application US/09839542B  
 ; Patent No. 6569996  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Blaschuk, Orest W.  
 ; APPLICANT: Symonds, James Matthew  
 ; APPLICANT: Gour, Barbara J.  
 ; TITLE OF INVENTION: COMPOUNDS AND METHODS FOR MODULATING NONCLASSICAL  
 ; TITLE OF INVENTION: CADHERIN-MEDIATED FUNCTIONS  
 ; FILE REFERENCE: 100086.407D1  
 ; CURRENT APPLICATION NUMBER: US/09/839,542B

; CURRENT FILING DATE: 2001-04-20  
 ; NUMBER OF SEQ ID NOS: 4052  
 ; SOFTWARE: PatentIn Ver. 2.0  
 ; SEQ ID NO 2953  
 ; LENGTH: 5  
 ; TYPE: PRT  
 ; ORGANISM: Artificial Sequence  
 ; FEATURE:  
 ; OTHER INFORMATION: Representative cyclic modulating agent based on  
 ; OTHER INFORMATION: Protocadherin cell adhesion recognition sequence  
 US-09-839-542B-2953

Query Match 0.7%; Score 5; DB 4; Length 5;  
 Best Local Similarity 100.0%; Pred. No. 3e+05;  
 Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 691 DDLVK 695  
 Db 1 DDLVK 5

RESULT 7  
 US-09-535-852-916  
 ; Sequence 916, Application US/09535852  
 ; Patent No. 6638911  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Blaschuk, Orest W.  
 ; APPLICANT: Symonds, James M.  
 ; APPLICANT: Gour, Barbara J.  
 ; TITLE OF INVENTION: COMPOUNDS AND METHODS FOR MODULATING  
 ; TITLE OF INVENTION: DESMOSOMAL CADHERIN-MEDIATED FUNCTIONS  
 ; FILE REFERENCE: 100086.407C6  
 ; CURRENT APPLICATION NUMBER: US/09/535,852  
 ; CURRENT FILING DATE: 2001-05-21  
 ; NUMBER OF SEQ ID NOS: 2009  
 ; SOFTWARE: FastSeq for Windows Version 4.0  
 ; SEQ ID NO 916  
 ; LENGTH: 5  
 ; TYPE: PRT  
 ; ORGANISM: Artificial Sequence  
 ; FEATURE:  
 ; OTHER INFORMATION: Cyclicized modulating agent comprising  
 ; OTHER INFORMATION: desmoglein-1 cell adhesion recognition sequence  
 US-09-535-852-916

Query Match 0.7%; Score 5; DB 4; Length 5;  
 Best Local Similarity 100.0%; Pred. No. 3e+05;  
 Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 242 CRALC 246  
 Db 1 CRALC 5

RESULT 8  
 US-07-609-716-7  
 ; Sequence 7, Application US/07609716  
 ; Patent No. 5514581  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Ferrari, Franco A.  
 ; APPLICANT: Cappello, Joseph  
 ; TITLE OF INVENTION: Functional Recombinantly Prepared  
 ; TITLE OF INVENTION: Synthetic Protein Polymer  
 ; NUMBER OF SEQUENCES: 118  
 ; CORRESPONDENCE ADDRESS:  
 ; ADDRESSEE: Flehr, Hobbach, Test, Albritton & Herbert  
 ; STREET: Four Embarcadero Center, Suite 3400  
 ; CITY: San Francisco  
 ; STATE: CA  
 ; COUNTRY: US  
 ; ZIP: 94111  
 ; COMPUTER READABLE FORM:  
 ; MEDIUM TYPE: Floppy disk

```

; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
;   APPLICATION NUMBER: US/07/609,716
;   FILING DATE: 06-NOV-1990
;   CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
;   NAME: Rowland, Bertram I.
;   REGISTRATION NUMBER: 20015
; REFERENCE/DOCKET NUMBER: A-55186-3/BIR
; TELEPHONE: 415-781-1989
; TELEFAX: 415-398-3249
; INFORMATION FOR SEQ ID NO: 7:
; SEQUENCE CHARACTERISTICS:
;   LENGTH: 6 amino acids
;   TYPE: amino acid
;   STRANDEDNESS: single
;   TOPOLOGY: linear
; MOLECULE TYPE: peptide
US-07-609-716-7

Query Match      0.7%; Score 5; DB 1; Length 6;
Best Local Similarity 100.0%; Pred. No. 3e+05;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      99 SGAGA 103
DB      1 SGAGA 5

RESULT 9
US-08-175-155-1
; Sequence 1, Application US/08175155
; Patent No. 5641648
; GENERAL INFORMATION:
; APPLICANT: Ferrari, Franco A.
; APPLICANT: Cappello, Joseph
; APPLICANT: Crisman, John W.
; APPLICANT: Dorman, Mary A.
; TITLE OF INVENTION: Methods for Preparing Synthetic
; TITLE OF INVENTION: Repetitive DNA
; NUMBER OF SEQUENCES: 69
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Flehr, Hombach, Test, Albritton & Herbert
; STREET: Four Embarcadero Center, Suite 3400
; CITY: San Francisco
; STATE: CA
; COUNTRY: US
; ZIP: 94111
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/175,155
; FILING DATE: 29-DEC-1993
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Rowland, Bertram I.
; REGISTRATION NUMBER: 20015
; REFERENCE/DOCKET NUMBER: A-55186-5/BIR
; TELEPHONE: 415-781-1989
; TELEFAX: 415-398-3249
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
;   LENGTH: 6 amino acids
;   TYPE: amino acid
;   STRANDEDNESS: single
;   TOPOLOGY: linear

```

```

; MOLECULE TYPE: peptide
US-08-175-155-1

Query Match      0.7%; Score 5; DB 1; Length 6;
Best Local Similarity 100.0%; Pred. No. 3e+05;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      99 SGAGA 103
DB      1 SGAGA 5

RESULT 10
US-08-469-005A-5
; Sequence 5, Application US/08469005A
; Patent No. 5665874
; GENERAL INFORMATION:
; APPLICANT: KUHAJDA, FRANCIS P.
; APPLICANT: PASTERNAK, GARY A.
; TITLE OF INVENTION: CANCER RELATED ANTIGEN
; NUMBER OF SEQUENCES: 14
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: BAKER & BOTTS, L.L.P.
; STREET: 1299 Pennsylvania Avenue, N.W.
; CITY: Washington
; STATE: DC
; COUNTRY: USA
; ZIP: 20004-2400
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSeq Version 1.5
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/469,005A
; FILING DATE: 05-JUN-1995
; CLASSIFICATION: 536
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/188,426
; FILING DATE: 24-JAN-1994
; APPLICATION NUMBER: 08/096,908
; FILING DATE: 26-JUL-1993
; APPLICATION NUMBER: 07/917,716
; FILING DATE: 24-JUL-1992
; ATTORNEY/AGENT INFORMATION:
; NAME: Posorske, Laurence H
; REGISTRATION NUMBER: 34,698
; REFERENCE/DOCKET NUMBER: 062482-0113
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 202-639-7700
; TELEFAX: 202-639-7890
; TELEX:
; INFORMATION FOR SEQ ID NO: 5:
; SEQUENCE CHARACTERISTICS:
;   LENGTH: 6 amino acids
;   TYPE: amino acid
;   STRANDEDNESS: single
;   TOPOLOGY: linear
; MOLECULE TYPE: peptide
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
; FRAGMENT TYPE: N-terminal
; ORIGINAL SOURCE:
; ORIGINAL SOURCE:
US-08-469-005A-5

Query Match      0.7%; Score 5; DB 1; Length 6;
Best Local Similarity 100.0%; Pred. No. 3e+05;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      230 AALQE 234
DB      2 AALQE 6

```

```

RESULT 11
US-08-211-747-4
; Sequence 4, Application US/08211747
; Patent No. 5733549
; GENERAL INFORMATION:
; APPLICANT: Yamada, Shingo
; APPLICANT: Inoue, Keiichi
; APPLICANT: Kitajima, Megumi
; APPLICANT: Yoshimura, Hajime
; APPLICANT: Sakurabayashi, Ikunosuke
; TITLE OF INVENTION: Peptides Including Amino Acid Sequences
; TITLE OF INVENTION: Selected From Lipoprotein (a) and Apolipoprotein (a),
; TITLE OF INVENTION: Antibodies Recognizing These Amino Acid Sequences, and
; TITLE OF INVENTION: Methods of Determination Using Antibodies
; NUMBER OF SEQUENCES: 10
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: PENNIE & EDMONDS
; STREET: 1155 Avenue of the Americas
; CITY: New York
; STATE: New York
; COUNTRY: U.S.A.
; ZIP: 10039-2711
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/211.747
; FILING DATE: 14-APR-1994
; CLASSIFICATION: 530
; ATTORNEY/AGENT INFORMATION:
; NAME: Mistock, S. Leslie
; REGISTRATION NUMBER: 18,872
; REFERENCE/DOCKET NUMBER: 7005-077
; TELEPHONE: (212) 790-9090
; TELEFAX: (212) 869-8864/9741
; TELEX: 66141 PENNIE
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 6 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; MOLECULE TYPE: peptide
; US-08-211-747-4

Query Match 0.7%; Score 5; DB 1; Length 6;
Best Local Similarity 100.0%; Pred. No. 3e+05;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 40 AEGTA 44
Db 2 AEGTA 6

RESULT 12
US-08-188-426-5
; Sequence 5, Application US/08188426
; Patent No. 5759791
; GENERAL INFORMATION:
; APPLICANT: Kuhajda, Francis P.
; APPLICANT: Pasternack, Gary R.
; TITLE OF INVENTION: CANCER RELATED ANTIGEN
; NUMBER OF SEQUENCES: 8
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Banner, Birch, McKie & Beckett
; STREET: 1001 G Street, N.W. -- 11th Floor
; CITY: Washington
; STATE: D.C.
; COUNTRY: U.S.A.

```

```

; ZIP: 20001
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/188.426
; FILING DATE: 24-JAN-1994
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Posorske, Laurence H.
; REGISTRATION NUMBER: 34,698
; REFERENCE/DOCKET NUMBER: 1107.45253
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 202 508-9100
; TELEFAX: 202 508-9299
; INFORMATION FOR SEQ ID NO: 5:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 6 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
; ORIGINAL SOURCE:
; ORGANISM: Homo sapiens
; US-08-188-426-5

Query Match 0.7%; Score 5; DB 1; Length 6;
Best Local Similarity 100.0%; Pred. No. 3e+05;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 230 AALQE 234
Db 2 AALQE 6

RESULT 13
US-08-477-509B-21
; Sequence 21, Application US/08477509B
; Patent No. 5770697
; GENERAL INFORMATION:
; APPLICANT: Ferrari, Franco A
; APPLICANT: Cappello, Joseph
; APPLICANT: Crissman, John W
; APPLICANT: Dorman, Mary A
; TITLE OF INVENTION: No. 5770697el Peptides Comprising Repetitive
; TITLE OF INVENTION: Units of Amino Acids and DNA Sequences Encoding the Same
; NUMBER OF SEQUENCES: 112
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Flehr, Hobbach, Test, Albritton & Herbert
; STREET: Four Embarcadero Center, Suite 3400
; CITY: San Francisco
; STATE: California
; COUNTRY: US
; ZIP: 94111
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/477.509B
; FILING DATE: 07-JUN-1995
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/175,155
; FILING DATE: 29-DEC-1993
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/053,049
; FILING DATE: 22-APR-1993
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/114,618

```

; FILING DATE: 29-OCT-1987  
 ; PRIOR APPLICATION DATA:  
 ; APPLICATION NUMBER: US 06/927,258  
 ; FILING DATE: 04-NOV-1986  
 ; ATTORNEY/AGENT INFORMATION:  
 ; NAME: Trecartin, Richard F.  
 ; REGISTRATION NUMBER: 31,801  
 ; REFERENCE/DOCKET NUMBER: A-55186-7/RFT/MTK  
 ; TELECOMMUNICATION INFORMATION:  
 ; TELEPHONE: 415-781-1989  
 ; TELEX: 910 277299  
 ; INFORMATION FOR SEQ ID NO: 21:  
 ; SEQUENCE CHARACTERISTICS:  
 ; LENGTH: 6 amino acids  
 ; TYPE: amino acid  
 ; STRANDEDNESS: single  
 ; TOPOLOGY: linear  
 ; MOLECULE TYPE: peptide  
 ; US-08-477-509B-21

Query Match 0.7%; Score 5; DB 1; Length 6;  
 Best Local Similarity 100.0%; Pred. No. 3e+05;  
 Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 99 SGAGA 103  
 |||||  
 Db 1 SGAGA 5

RESULT 14  
 US-08-707-237A-6  
 ; Sequence 6, Application US/08707237A  
 ; Patent No. 5820713  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Ferrari, Franco A.  
 ; APPLICANT: Capello, Joseph  
 ; APPLICANT: Criseman, John W.  
 ; APPLICANT: Dorman, Mary A.  
 ; TITLE OF INVENTION: METHODS FOR PREPARING SYNTHETIC  
 ; TITLE OF INVENTION: REPETITIVE DNA  
 ; NUMBER OF SEQUENCES: 108  
 ; CORRESPONDENCE ADDRESS:  
 ; ADDRESSEE: Flehr, Hohnbach, Test, Albritton & Herbert  
 ; STREET: Four Embarcadero Center, Suite 3400  
 ; CITY: San Francisco  
 ; STATE: California  
 ; COUNTRY: United States  
 ; ZIP: 94111-4187  
 ; COMPUTER READABLE FORM:  
 ; MEDIUM TYPE: Floppy disk  
 ; COMPUTER: IBM PC compatible  
 ; OPERATING SYSTEM: PC-DOS/MS-DOS  
 ; SOFTWARE: PatentIn Release #1.0, Version #1.30  
 ; CURRENT APPLICATION DATA:  
 ; APPLICATION NUMBER: US/08/707,237A  
 ; FILING DATE: 03-SEP-1996  
 ; CLASSIFICATION: 435  
 ; PRIOR APPLICATION DATA:  
 ; APPLICATION NUMBER: US 08/175,155  
 ; FILING DATE: 29-DEC-1993  
 ; PRIOR APPLICATION DATA:  
 ; APPLICATION NUMBER: US 08/053,049  
 ; FILING DATE: 22-APR-1993  
 ; PRIOR APPLICATION DATA:  
 ; APPLICATION NUMBER: US 07/609,716  
 ; FILING DATE: 06-NOV-1990  
 ; PRIOR APPLICATION DATA:  
 ; APPLICATION NUMBER: US 07/269,429  
 ; FILING DATE: 09-NOV-1988  
 ; PRIOR APPLICATION DATA:  
 ; APPLICATION NUMBER: US 07/114,618  
 ; FILING DATE: 29-OCT-1987  
 ; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: US 06/927,258  
 ; FILING DATE: 04-NOV-1986  
 ; ATTORNEY/AGENT INFORMATION:  
 ; NAME: Trecartin, Richard F.  
 ; REGISTRATION NUMBER: 31,801  
 ; REFERENCE/DOCKET NUMBER: A-55186-10/WHD  
 ; TELECOMMUNICATION INFORMATION:  
 ; TELEPHONE: (415) 781-1989  
 ; TELEX: 910 277299  
 ; INFORMATION FOR SEQ ID NO: 6:  
 ; SEQUENCE CHARACTERISTICS:  
 ; LENGTH: 6 amino acids  
 ; TYPE: amino acid  
 ; STRANDEDNESS: unknown  
 ; TOPOLOGY: unknown  
 ; MOLECULE TYPE: protein  
 ; US-08-707-237A-6

Query Match 0.7%; Score 5; DB 2; Length 6;  
 Best Local Similarity 100.0%; Pred. No. 3e+05;  
 Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 99 SGAGA 103  
 |||||  
 Db 1 SGAGA 5

RESULT 15  
 US-08-659-486A-4  
 ; Sequence 4, Application US/08659486A  
 ; Patent No. 5840499  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Lawrence Brass and James A. Hoxie  
 ; TITLE OF INVENTION: Method and Kit for the Detection of  
 ; TITLE OF INVENTION: Thrombin Receptor Activation of Platelet and other Cells  
 ; NUMBER OF SEQUENCES: 4  
 ; CORRESPONDENCE ADDRESS:  
 ; ADDRESSEE: Jane Massey Licata, Esq.  
 ; STREET: 210 Lake Drive East, Suite 201  
 ; CITY: Cherry Hill  
 ; STATE: NJ  
 ; COUNTRY: USA  
 ; ZIP: 08002  
 ; COMPUTER READABLE FORM:  
 ; MEDIUM TYPE: DISKETTE, 3.5 INCH, 1.44 MB STORAGE  
 ; COMPUTER: IBM 486  
 ; OPERATING SYSTEM: WINDOWS FOR WORKGROUPS  
 ; SOFTWARE: WORDPERFECT 5.1  
 ; CURRENT APPLICATION DATA:  
 ; APPLICATION NUMBER: US/08/659,486A  
 ; FILING DATE: June 6, 1996  
 ; CLASSIFICATION: 435  
 ; PRIOR APPLICATION DATA:  
 ; APPLICATION NUMBER: 08/220,375  
 ; FILING DATE: March 31, 1994  
 ; ATTORNEY/AGENT INFORMATION:  
 ; NAME: Jane Massey Licata  
 ; REGISTRATION NUMBER: 32,257  
 ; REFERENCE/DOCKET NUMBER: PENN-0331  
 ; TELECOMMUNICATION INFORMATION:  
 ; TELEPHONE: (609) 779-2400  
 ; TELEFAX: (609) 779-8488  
 ; INFORMATION FOR SEQ ID NO: 4:  
 ; SEQUENCE CHARACTERISTICS:  
 ; LENGTH: 6  
 ; TYPE: amino acid  
 ; TOPOLOGY: linear  
 ; US-08-659-486A-4

Query Match 0.7%; Score 5; DB 2; Length 6;  
 Best Local Similarity 100.0%; Pred. No. 3e+05;  
 Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 677 TLDP 681  
| | | | |  
Db 2 TLDP 6

Search completed: August 17, 2004, 10:17:38  
Job time : 20 secs

This Page Blank (uspio)

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: July 18, 2004, 09:45:24 ; Search time 24 Seconds  
(without alignments)  
2949.873 Million cell updates/sec

Title: US-10-045-815-4  
Perfect score: 736  
Sequence: 1 MAVRALKLLTTLAVVAAS.....PPEPAQESLSGSESKPKDEL 736

Scoring table: OLIGO  
Gapop 60.0 , Gapext 60.0

Searched: 293366 seqs, 96191526 residues

Word size : 0

Total number of hits satisfying chosen parameters: 283366

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database : PIR 78:\*  
1: pir1:\*  
2: pir2:\*  
3: pir3:\*  
4: pir4:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	9	1.2	364	1 A55014	GTP-binding protei
2	8	1.1	222	2 H81354	probable undecapre
3	8	1.1	257	3 T32002	hypothetical prote
4	8	1.1	317	2 T35010	probable integral
5	8	1.1	431	2 A56922	synaptonemal compl
6	8	1.1	600	2 T41165	hypothetical prote
7	8	1.1	623	2 AB1112	transcription anti
8	8	1.1	623	2 AF1473	transcription anti
9	8	1.1	913	2 T51923	related to heterok
10	8	1.1	973	2 T01862	hypothetical prote
11	8	1.1	1233	1 G71612	novel serine/threo
12	8	1.1	1384	2 T02748	hypothetical prote
13	8	1.1	1384	2 T02748	hypothetical prote
14	7	1.0	2895	2 H85362	hypothetical prote
15	7	1.0	34	2 F82163	hypothetical prote
16	7	1.0	54	2 C84292	hypothetical prote
17	7	1.0	73	2 AC0914	probable lipoprote
18	7	1.0	76	2 B71840	hypothetical prote
19	7	1.0	76	2 B64675	conserved hypotet
20	7	1.0	81	2 T03711	reverse transcript
21	7	1.0	82	2 A61040	transforming growt
22	7	1.0	89	2 S72742	B117 Fl 32 protei
23	7	1.0	103	2 S37966	hypothetical prote
24	7	1.0	107	2 T23910	hypothetical prote
25	7	1.0	110	2 E70814	hypothetical prote
26	7	1.0	121	2 I58134	transforming growt
27	7	1.0	123	2 B30560	Ig heavy chain V r
28	7	1.0	129	2 AD0029	50S ribosomal prot
29	7	1.0	140	2 G69813	transcription regu
30	7	1.0	142	2 E95176	acetyltransferase,

#### ALIGNMENTS

##### RESULT 1

A55014  
GTP-binding protein DRG homolog - human  
C:Species: Homo sapiens (man)  
C:Date: 24-Jul-1998 #sequence\_revision 24-Jul-1998 #text\_change 19-Jan-2001  
R;Schenker, T.; Lach, C.; Kessler, B.; Calderara, S.; Trueb, B.  
J. Biol. Chem. 269, 25447-25453, 1994  
A:Title: A novel GTP-binding protein which is selectively repressed in SV40 transformed f  
A:Reference number: A55014; MUID:95014343; PMID:7929244  
A:Accession: A55014  
A>Status: preliminary  
A:Molecule type: mRNA  
A:Residues: 1-364 <SCH>  
A:Cross-references: GB:X80754; NID:G577778; PIDN:CAA56730.1; PID:G577779  
A:Experimental source: fibroblasts  
C:Comment: This protein is expressed in normal fibroblasts but repressed in transformed c  
C:Genetics: GDB:NEDD3; DRG1

30	7	1.0	145	2	C87674	hypothetical prote
31	7	1.0	157	2	H83543	hypothetical prote
32	7	1.0	159	1	WPRT1	transforming growt
33	7	1.0	159	1	S27195	transforming growt
34	7	1.0	159	2	IS7497	transforming growt
35	7	1.0	160	1	WRHUI	transforming growt
36	7	1.0	160	2	S39795	transforming growt
37	7	1.0	160	2	G98042	hypothetical prote
38	7	1.0	171	2	B75446	conserved hypotet
39	7	1.0	177	2	F97674	hypothetical prote
40	7	1.0	177	2	A32899	conserved hypotet
41	7	1.0	180	2	D90318	coenzyme F390 synt
42	7	1.0	188	2	I79500	myc protein - huma
43	7	1.0	188	2	JU0451	hypothetical 21K p
44	7	1.0	188	2	I59116	myc protein - huma
45	7	1.0	188	2	A72745	hypothetical prote

##### RESULT 2

H81354  
Probable undecaprenyl diphosphate synthase Cj0824 [imported] - Campylobacter jejuni (str  
C:Species: Campylobacter jejuni  
C:Date: 31-Mar-2000 #sequence\_revision 31-Mar-2000 #text\_change 03-Jun-2002  
C:Accession: H81354  
R;Parkhill, J.; Wren, B.W.; Mungall, K.; Kettle, J.M.; Churcher, C.; Basham, D.; Chilling  
C.W.; Quail, M.; Rajandream, M.A.; Rutherford, K.M.; VanVliet, A.; Whithead, S.; Barrell  
Nature 403, 665-668, 2000  
A:Title: The genome sequence of the food-borne pathogen Campylobacter jejuni reveals hyp  
A:Reference number: A81250; MUID:20150912; PMID:10688204  
A:Accession: H81354  
A>Status: preliminary  
A:Molecule type: DNA

Query Match 1.2%; Score 9; DB 1; Length 364;  
Best Local Similarity 100.0%; Pred. No. 1.6;  
Matches 367 QRSLLLEKEL 375  
Db 1 155 QRSLLLEKEL 163

Mismatches 0; Indels 0; Gaps 0;





A;Status: preliminary  
A;Molecule type: DNA  
A;Residues: 1-623 <GLA>  
A;Cross-references: GB:NC\_003210; PIDN:CAD00824.1; PID:g16409662; GSPDB:GN00177  
A;Experimental source: strain EGD-e  
C;Genetics:  
A;Gene: lmo0297

Query Match 1.1%; Score 8; DB 2; Length 623;  
Best Local Similarity 100.0%; Pred. No. 27;  
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 366 RORSLEK 373  
Db 6 RORSLEK 13  
|||||

RESULT 8  
AF1473  
transcription antiterminator (BglG family) homolog lin0325 [imported] - *Listeria innocua*  
C;Species: *Listeria innocua*  
C;Date: 27-Nov-2001 #sequence\_revision 27-Nov-2001 #text\_change 27-Nov-2001  
C;Accession: AF1473  
R;Glaser, P.; Frangeul, L.; Buchrieser, C.; Amend, A.; Baquero, F.; Berche, P.; Bloeker, D.; Dominguez-Bernal, G.; Duchaud, E.; Durand, L.; Dussurget, O.; Entian, K.D.; Fsihi, H.; Jones, L.M.; Karst, U.  
Science 294, 849-852, 2001  
A;Authors: Kraft, J.; Kuhn, M.; Kunst, F.; Kurapkat, G.; Madueno, E.; Maitournam, A.; Mak, C.; Schluter, T.; Simoes, N.; Tierrez, A.; Vazquez-Boland, J.A.; Voss, H.; Wehland, A.; Title: Comparative genomics of *Listeria* species.  
A;Reference number: AB1077; MUID:21537279; PMID:11679669  
A;Accession: AF1473  
A;Status: preliminary  
A;Molecule type: DNA  
A;Residues: 1-623 <GLA>  
A;Cross-references: GB:AL592022; PIDN:CAC95558.1; PID:g16412754; GSPDB:GN00178  
A;Experimental source: strain clipl1262  
C;Genetics:  
A;Gene: lin0325

Query Match 1.1%; Score 8; DB 2; Length 623;  
Best Local Similarity 100.0%; Pred. No. 27;  
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 366 RORSLEK 373  
Db 6 RORSLEK 13  
|||||

RESULT 9  
T51923  
related to heterokaryon incompatibility protein het-6 [imported] - *Neurospora crassa*  
N;Alternate names: protein B2311.350  
C;Species: *Neurospora crassa*  
C;Date: 20-Oct-2000 #sequence\_revision 20-Oct-2000 #text\_change 20-Oct-2000  
C;Accession: T51923  
R;Schulte, U.; Aign, V.; Hoheisel, J.; Brandt, P.; Fartmann, B.; Holland, R.; Nyakatura, submitted to the Protein Sequence Database, August 2000  
A;Reference number: 225958  
A;Accession: T51923  
A;Status: preliminary  
A;Molecule type: DNA  
A;Residues: 1-913 <SCH>  
A;Cross-references: EMBL:AL391572; GSPDB:GN00116; NCSP:B23111.350  
A;Experimental source: BAC clone B23111; strain OR74A  
C;Genetics:  
A;Gene: NCSP:B23111.350  
A;Map position: 6  
A;Introns: 13/1; 410/2; 516/2; 867/3; 887/3

Query Match 1.1%; Score 8; DB 2; Length 913;  
Best Local Similarity 100.0%; Pred. No. 38;  
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 66 AALRALRL 73  
Db 213 AALRALRL 220  
|||||

RESULT 10  
T01862  
hypothetical protein T7M24.4 - *Arabidopsis thaliana*  
C;Species: *Arabidopsis thaliana* (mouse-ear cress)  
C;Date: 26-Feb-1999 #sequence\_revision 26-Feb-1999 #text\_change 24-Mar-1999  
C;Accession: T01862  
R;Harmon, G.; Langston, Y.; Stoneking, T.; Drone, K.; Ames, M.  
submitted to the EMBL Data Library, July 1998  
A;Description: The sequence of *Arabidopsis thaliana* T7M24.  
A;Reference number: Z1448  
A;Accession: T01862  
A;Status: translated from GB/EMBL/DDBJ  
A;Molecule type: DNA  
A;Residues: 1-973 <HAR>  
A;Cross-references: EMBL:AF077408; NID:g3319359; PID:g3319363  
A;Experimental source: cultivar Columbia  
C;Genetics:  
A;Map position: 4  
A;Introns: 257/1; 284/1; 378/3; 689/3; 744/3  
A;Note: T7M24.4

Query Match 1.1%; Score 8; DB 2; Length 973;  
Best Local Similarity 100.0%; Pred. No. 40;  
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 429 ETLVBEKT 436  
Db 19 ETLVBEKT 26  
|||||

RESULT 11  
G71612  
novel serine/threonine-specific protein kinase (EC 2.7.1.1-) PF0520W - malaria parasite  
C;Species: *Plasmodium falciparum*  
C;Date: 10-Sep-1999 #sequence\_revision 10-Sep-1999 #text\_change 09-Jun-2000  
C;Accession: G71612  
R;Gardner, M.J.; Tettelin, H.; Carucci, D.J.; Cummings, L.M.; Aravind, L.; Koonin, E.V.; Pertea, M.; Salzberg, S.; Zhou, L.; Sutton, G.G.; Clayton, R.; White, O.; Smith, H.O.; Science 282, 1126-1132, 1998  
A;Title: Chromosome 2 sequence of the human malaria parasite *Plasmodium falciparum*.  
A;Reference number: AV1600; MUID:99021743; PMID:9804551  
A;Accession: G71612  
A;Status: preliminary; nucleic acid sequence not shown; translation not shown  
A;Molecule type: DNA  
A;Residues: 1-1233 <GAR>  
A;Cross-references: GB:AE001400; GB:AE001362; NID:g3845204; PIDN:AACT1893.1; PID:g3845205  
A;Experimental source: clone 3D7  
C;Genetics:  
A;Gene: PF0520W  
C;Superfamily: malaria parasite serine/threonine-specific protein kinase PF0520W; protei  
C;Keywords: phosphotransferase  
F;949-1221/Domain: protein kinase homology <KIN>

Query Match 1.1%; Score 8; DB 1; Length 1233;  
Best Local Similarity 100.0%; Pred. No. 49;  
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 154 FKINKLEK 161  
Db 771 FKINKLEK 778  
|||||

RESULT 12  
T02748  
hypothetical protein R31180.1 - human  
C;Species: *Homo sapiens* (man)  
C;Date: 02-Sep-2000 #sequence\_revision 02-Sep-2000 #text\_change 31-Dec-2000

C:Accession: T02748  
 R:lamerdin, J.E.; McCready, P.M.; Skowronski, E.; Adamson, A.W.; Burkhardt-Schultz, K.; G  
 J.; Danganan, L.; Poundstone, P.; Christensen, M.; Georgescu, A.; Avila, J.; Liu, S.; Ab  
 P.; Quan, G.; Krommiller, B.; Arellano, A.; Montgomery, M.; Ow, D.; Nolan, M.  
 Submitted to the EMBL Data Library, August 1998  
 A:Authors: Trong, S.; Kobayashi, A.; Olsen, A.S.; Carrano, A.V.  
 A:Description: Sequence analysis of a 3.5 Mb contig in human 19p13.3 containing a serine  
 A:Accession number: Z14714  
 A:Accession: T02748  
 A:Status: preliminary; translated from GB/EMBL/DBJ  
 A:Molecule type: DNA  
 A:Residues: 1-1384 <LAM>  
 A:Cross-references: EMBL:AC005390; NID:g3399675; PIDN:AAC28919.1; PID:g3399676  
 C:Genetics:  
 A:Map position: 19  
 A:Map position: 19  
 A:Introns: 15/3; 51/1; 131/2; 180/3; 216/1; 266/3; 311/2; 359/3; 407/3; 447/1; 482/2; 53  
 3; 1160/2; 1224/1  
 C:Supertfamily: Arabidopsis thaliana hypothetical protein YUP8H12R.3

Query Match 1.1%; Score 8; DB 2; Length 1384;  
 Best Local Similarity 100.0%; Pred. No. 55;  
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 296 FEDFLPSH 303  
 |||||  
 DB 164 FEDFLPSH 171

RESULT 13  
 H85362  
 Hypothetical protein AT4G30990 [imported] - Arabidopsis thaliana  
 C:Species: Arabidopsis thaliana (mouse-ear cress)  
 C:Date: 16-Feb-2001 #sequence\_revision 16-Feb-2001 #text\_change 16-Feb-2001  
 C:Accession: H85362  
 R:Anonymous, The European Union Arabidopsis Genome Sequencing Consortium, The Cold Spring  
 Nature 402, 769-777, 1999  
 A:Title: Sequence and analysis of chromosome 4 of the plant Arabidopsis thaliana.  
 A:Reference number: A85001; MUID:20083488; PMID:10617198  
 A:Accession: H85362  
 A:Status: preliminary  
 A:Molecule type: DNA  
 A:Residues: 1-2895 <STO>  
 A:Cross-references: GB:NC\_001269; NID:g7270001; PIDN:CAB79817.1; GSPDB:GN00140  
 A:Gene: AT4G30990  
 A:Map position: 4

Query Match 1.1%; Score 8; DB 2; Length 2895;  
 Best Local Similarity 100.0%; Pred. No. 1.1e+02;  
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 6 LKLLTLL 13  
 |||||  
 DB 1916 LKLLTLL 1923

RESULT 14  
 F82163  
 Hypothetical protein VC1734 [imported] - Vibrio cholerae (strain N16961 serogroup O1)  
 C:Species: Vibrio cholerae  
 C:Date: 18-Aug-2000 #sequence\_revision 20-Aug-2000 #text\_change 02-Feb-2001  
 C:Accession: F82163  
 R:Heidelberg, J.F.; Eisen, J.A.; Nelson, W.C.; Clayton, R.A.; Gwinn, M.L.; Dodson, R.J.;  
 chardson, D.; Ermolaeva, M.D.; Vamathevan, J.; Bass, S.; Qin, H.; Dragoi, I.; Sellers, F  
 1, R.R.; Mekalanos, J.J.; Venter, J.C.; Fraser, C.M.  
 Nature 406, 477-483, 2000  
 A:Title: DNA Sequence of both chromosomes of the cholera pathogen Vibrio cholerae.  
 A:Reference number: A82035; MUID:20406833; PMID:10952301  
 A:Accession: F82163  
 A:Status: preliminary  
 A:Molecule type: DNA  
 A:Residues: 1-34 <HEI>  
 A:Cross-references: GB:AE004251; GB:AE003852; NID:g956248; PIDN:AAF94884.1; GSPDB:GN001

A:Experimental source: serogroup O1; strain N16961; biotype El Tor  
 C:Genetics:  
 A:Gene: VC1734  
 A:Map position: 1

Query Match 1.0%; Score 7; DB 2; Length 34;  
 Best Local Similarity 100.0%; Pred. No. 20;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 128 AAHSLSE 134  
 |||||  
 DB 11 AAHSLSE 17

RESULT 15  
 C84292  
 Hypothetical protein Vng1376h [imported] - Halobacterium sp. NRC-1  
 C:Species: Halobacterium sp. NRC-1  
 C:Date: 02-Feb-2001 #sequence\_revision 02-Feb-2001 #text\_change 02-Feb-2001  
 C:Accession: C84292  
 R:Ng, W.V.; Kennedy, S.P.; Mahairas, G.G.; Berquist, B.; Pan, M.; Shukla, H.D.; Lasky, S.  
 ; Leitthausen, B.; Keller, K.; Cruz, R.; Danson, M.J.; Hough, D.W.; Maddocks, D.G.; Jablon  
 Jung, K.H.; Alam, M.; Freitas, T.  
 Proc. Natl. Acad. Sci. U.S.A. 97, 12176-12181, 2000  
 A:Authors: Hou, S.; Daniels, C.J.; Dennis, P.P.; Omer, A.D.; Ebhardt, H.; Lowe, T.M.; Lia  
 A:Title: Genome sequence of Halobacterium species NRC-1.  
 A:Reference number: A84160; MUID:20504483; PMID:11016950  
 A:Accession: C84292  
 A:Status: preliminary  
 A:Molecule type: DNA  
 A:Residues: 1-54 <STO>  
 A:Cross-references: GB:AE004437; NID:gi0580885; PIDN:AAG19703.1; GSPDB:GN00138  
 C:Genetics:  
 A:Gene: VNG1376H

Query Match 1.0%; Score 7; DB 2; Length 54;  
 Best Local Similarity 100.0%; Pred. No. 31;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 227 HLEALQ 233  
 |||||  
 DB 11 HLEALQ 17

Search completed: July 18, 2004, 10:00:23  
 Job time : 26 secs



QY 731 KPKDEL 736  
| | | | |  
DB 1 KPKDEL 6

## RESULT 2

US-10-367-593-129  
; Sequence 129, Application US/10367593  
; Publication No. US20040071721A1  
; GENERAL INFORMATION:  
; APPLICANT: Rothman, James E.  
; APPLICANT: Hartl, F. Ulrich  
; APPLICANT: Hoe, Mee H.  
; APPLICANT: Houghton, Alan  
; APPLICANT: Takechi, Yoshizumi  
; APPLICANT: Mayhew, Mark  
; TITLE OF INVENTION: Heat Shock Protein-Based Vaccines and Immunotherapies  
; FILE REFERENCE: 11746/461012  
; CURRENT APPLICATION NUMBER: US/10/367,593  
; CURRENT FILING DATE: 2003-02-14  
; PRIOR APPLICATION NUMBER: US 09/011,645  
; PRIOR FILING DATE: 1998-02-13  
; PRIOR APPLICATION NUMBER: PCT/US96/13363  
; PRIOR FILING DATE: 1996-08-16  
; PRIOR APPLICATION NUMBER: US 60/002,490  
; PRIOR FILING DATE: 1995-08-18  
; PRIOR APPLICATION NUMBER: US 60/002,479  
; PRIOR FILING DATE: 1995-08-18  
; NUMBER OF SEQ ID NOS: 349  
; SOFTWARE: WordPerfect 8.0 for Windows  
; SEQ ID NO 129  
; LENGTH: 8  
; TYPE: PRT  
; ORGANISM: Artificial Sequence  
; FEATURE:  
; OTHER INFORMATION: synthetic peptide  
US-10-367-593-129

Query Match 0.8%; Score 6; DB 12; Length 8;  
Best Local Similarity 100.0%; Pred. No. 1.2e+06;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 731 KPKDEL 736  
| | | | |  
DB 1 KPKDEL 6

## RESULT 3

US-10-367-594-129  
; Sequence 129, Application US/10367594  
; Publication No. US20040071722A1  
; GENERAL INFORMATION:  
; APPLICANT: Rothman, James E.  
; APPLICANT: Hartl, F. Ulrich  
; APPLICANT: Hoe, Mee H.  
; APPLICANT: Houghton, Alan  
; APPLICANT: Takechi, Yoshizumi  
; APPLICANT: Mayhew, Mark  
; TITLE OF INVENTION: Heat Shock Protein-Based Vaccines and Immunotherapies  
; FILE REFERENCE: 11746/461041  
; CURRENT APPLICATION NUMBER: US/10/367,594  
; CURRENT FILING DATE: 2003-02-14  
; PRIOR APPLICATION NUMBER: US 09/680,806  
; PRIOR FILING DATE: 2000-10-05  
; PRIOR APPLICATION NUMBER: US 09/011,645  
; PRIOR FILING DATE: 1998-02-13  
; PRIOR APPLICATION NUMBER: PCT/US96/13363  
; PRIOR FILING DATE: 1996-08-16  
; PRIOR APPLICATION NUMBER: US 60/002,490  
; PRIOR FILING DATE: 1995-08-18  
; PRIOR APPLICATION NUMBER: US 60/002,479  
; PRIOR FILING DATE: 1995-08-18

; NUMBER OF SEQ ID NOS: 349  
; SOFTWARE: WordPerfect 8.0 for Windows  
; SEQ ID NO 129  
; LENGTH: 8  
; TYPE: PRT  
; ORGANISM: Artificial Sequence  
; FEATURE:  
; OTHER INFORMATION: synthetic peptide  
US-10-367-594-129

Query Match 0.8%; Score 6; DB 12; Length 8;  
Best Local Similarity 100.0%; Pred. No. 1.2e+06;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 731 KPKDEL 736  
| | | | |  
DB 1 KPKDEL 6

## RESULT 4

US-10-367-654-129  
; Sequence 129, Application US/10367654  
; Publication No. US20040071723A1  
; GENERAL INFORMATION:  
; APPLICANT: Rothman, James E.  
; APPLICANT: Hartl, F. Ulrich  
; APPLICANT: Hoe, Mee H.  
; APPLICANT: Houghton, Alan  
; APPLICANT: Takechi, Yoshizumi  
; APPLICANT: Mayhew, Mark  
; TITLE OF INVENTION: Heat Shock Protein-Based Vaccines and Immunotherapies  
; FILE REFERENCE: 11746/461032  
; CURRENT APPLICATION NUMBER: US/10/367,654  
; CURRENT FILING DATE: 2003-02-14  
; PRIOR APPLICATION NUMBER: US 10/171,734  
; PRIOR FILING DATE: 2002-06-13  
; PRIOR APPLICATION NUMBER: US 09/636,295  
; PRIOR FILING DATE: 2000-08-10  
; PRIOR APPLICATION NUMBER: US 09/011,645  
; PRIOR FILING DATE: 1998-02-13  
; PRIOR APPLICATION NUMBER: PCT/US96/13363  
; PRIOR FILING DATE: 1996-08-16  
; PRIOR APPLICATION NUMBER: US 60/002,490  
; PRIOR FILING DATE: 1995-08-18  
; PRIOR APPLICATION NUMBER: US 60/002,479  
; PRIOR FILING DATE: 1995-08-18  
; NUMBER OF SEQ ID NOS: 349  
; SOFTWARE: WordPerfect 8.0 for Windows  
; SEQ ID NO 129  
; LENGTH: 8  
; TYPE: PRT  
; ORGANISM: Artificial Sequence  
; FEATURE:  
; OTHER INFORMATION: synthetic peptide  
US-10-367-654-129

Query Match 0.8%; Score 6; DB 12; Length 8;  
Best Local Similarity 100.0%; Pred. No. 1.2e+06;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 731 KPKDEL 736  
| | | | |  
DB 1 KPKDEL 6

## RESULT 5

US-10-367-658-129  
; Sequence 129, Application US/10367658  
; Publication No. US20040071724A1  
; GENERAL INFORMATION:  
; APPLICANT: Rothman, James E.  
; APPLICANT: Hartl, F. Ulrich  
; APPLICANT: Hoe, Mee H.

APPLICANT: Houghton, Alan  
APPLICANT: Takechi, Yoshizumi  
APPLICANT: Mayhew, Mark  
TITLE OF INVENTION: Heat Shock Protein-Based Vaccines and Immunotherapies  
FILE REFERENCE: 11746/461051  
CURRENT APPLICATION NUMBER: US/10/367,658  
CURRENT FILING DATE: 2003-02-14  
PRIOR APPLICATION NUMBER: US 09/794,529  
PRIOR FILING DATE: 2001-02-27  
PRIOR APPLICATION NUMBER: US 09/011,645  
PRIOR FILING DATE: 1998-02-13  
PRIOR APPLICATION NUMBER: PCT/US96/13363  
PRIOR FILING DATE: 1996-08-16  
PRIOR APPLICATION NUMBER: US 60/002,490  
PRIOR FILING DATE: 1995-08-18  
PRIOR APPLICATION NUMBER: US 60/002,479  
PRIOR FILING DATE: 1995-08-18  
NUMBER OF SEQ ID NOS: 349  
SOFTWARE: WordPerfect 8.0 for Windows  
SEQ ID NO 129  
LENGTH: 8  
TYPE: PRT  
ORGANISM: Artificial Sequence  
FEATURE:  
OTHER INFORMATION: synthetic peptide  
US-10-367-658-129

Query Match 0.8%; Score 6; DB 12; Length 8;  
Best Local Similarity 100.0%; Pred. No. 1.2e+06;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
Qy 731 KPDEL 736  
Db 1 KPDEL 6

RESULT 6  
US-10-367-668-129  
Sequence 129, Application US/10367668  
Publication No. US20040071725A1  
GENERAL INFORMATION:  
APPLICANT: Rothman, James E.  
APPLICANT: Hartl, F. Ulrich  
APPLICANT: Hoe, Mee H.  
APPLICANT: Houghton, Alan  
APPLICANT: Takechi, Yoshizumi  
APPLICANT: Mayhew, Mark  
TITLE OF INVENTION: Heat Shock Protein-Based Vaccines and Immunotherapies  
FILE REFERENCE: 11746/461072  
CURRENT APPLICATION NUMBER: US/10/367,668  
CURRENT FILING DATE: 2003-02-14  
PRIOR APPLICATION NUMBER: US 09/794,517  
PRIOR FILING DATE: 2001-02-27  
PRIOR APPLICATION NUMBER: US 09/011,645  
PRIOR FILING DATE: 1998-02-13  
PRIOR APPLICATION NUMBER: PCT/US96/13363  
PRIOR FILING DATE: 1996-08-16  
PRIOR APPLICATION NUMBER: US 60/002,490  
PRIOR FILING DATE: 1995-08-18  
PRIOR APPLICATION NUMBER: US 60/002,479  
PRIOR FILING DATE: 1995-08-18  
NUMBER OF SEQ ID NOS: 349  
SOFTWARE: WordPerfect 8.0 for Windows  
SEQ ID NO 129  
LENGTH: 8  
TYPE: PRT  
ORGANISM: Artificial Sequence  
FEATURE:  
OTHER INFORMATION: synthetic peptide  
US-10-367-668-129

Query Match 0.8%; Score 6; DB 12; Length 8;  
Best Local Similarity 100.0%; Pred. No. 1.2e+06;

Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
Qy 731 KPDEL 736  
Db 1 KPDEL 6  
RESULT 7  
US-10-128-711-17  
Sequence 17, Application US/10128711  
Publication No. US20030099634A1  
GENERAL INFORMATION:  
APPLICANT: VITIELLO, Maria A.  
APPLICANT: CHESNUT, Robert W.  
APPLICANT: SETTE, Alessandro D.  
APPLICANT: CELIS, Esteban  
APPLICANT: GRAY, Howard  
TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR ELICITING CTL IMMUNITY  
NUMBER OF SEQUENCES: 153  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Townsend and Townsend Kourie and Crew  
STREET: Steuart Street Tower, One Market Plaza  
CITY: San Francisco  
STATE: California  
COUNTRY: US  
ZIP: 94105-1493  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/10/128,711  
FILING DATE: 22-Apr-2002  
CLASSIFICATION: <Unknown>  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US/08/197,484  
FILING DATE: 16-FEB-1994  
APPLICATION NUMBER: US 07/935,811  
FILING DATE: 26-AUG-1992  
APPLICATION NUMBER: US 07/874,491  
FILING DATE: 27-APR-1992  
APPLICATION NUMBER: US 07/827,682  
FILING DATE: 29-JAN-1992  
APPLICATION NUMBER: US 07/749,568  
FILING DATE: 26-AUG-1991  
ATTORNEY/AGENT INFORMATION:  
NAME: Parmelee, Steven W.  
REGISTRATION NUMBER: 31,990  
REFERENCE/DOCKET NUMBER: 14137-26-4  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (206) 467-9600  
TELEFAX: (206) 623-6793  
INFORMATION FOR SEQ ID NO: 17:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 8 amino acids  
TYPE: amino acid  
STRANDEDNESS: unknown  
TOPOLOGY: unknown  
MOLECULE TYPE: peptide  
SEQUENCE DESCRIPTION: SEQ ID NO: 17:  
US-10-128-711-17

Query Match 0.8%; Score 6; DB 14; Length 8;  
Best Local Similarity 100.0%; Pred. No. 1.2e+06;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
Qy 731 KPDEL 736  
Db 1 KPDEL 6

RESULT 8  
 US-10-079-167-4  
 ; Sequence 4, Application US/10079167  
 ; Publication No. US20030138454A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Hill, Adrian V.S.  
 ; APPLICANT: McShane, Helen  
 ; APPLICANT: Gilbert, Sarah C.  
 ; APPLICANT: Reece, William  
 ; APPLICANT: Schneider, Joerg  
 ; TITLE OF INVENTION: Vaccination Method  
 ; FILE REFERENCE: 2907.1000-001  
 ; CURRENT APPLICATION NUMBER: US/10/079,167  
 ; CURRENT FILING DATE: 2002-02-19  
 ; PRIOR APPLICATION NUMBER: US 09/454,204  
 ; PRIOR FILING DATE: 1999-12-09  
 ; PRIOR APPLICATION NUMBER: PCT/GB98/01681  
 ; PRIOR FILING DATE: 1998-06-09  
 ; PRIOR APPLICATION NUMBER: GB 97 11957.2  
 ; PRIOR FILING DATE: 1997-06-09  
 ; PRIOR APPLICATION NUMBER: PCT/GB01/04116  
 ; PRIOR FILING DATE: 2001-09-13  
 ; PRIOR APPLICATION NUMBER: GB 00 23203.3  
 ; PRIOR FILING DATE: 2001-09-21  
 ; NUMBER OF SEQ ID NOS: 99  
 ; SOFTWARE: FastSeq for Windows Version 4.0  
 ; SEQ ID NO 4  
 ; LENGTH: 8  
 ; TYPE: PRT  
 ; ORGANISM: Unknown  
 ; FEATURE:  
 ; OTHER INFORMATION: CTL Epitope of the Malaria String  
 US-10-079-167-4

Query Match 0.8%; Score 6; DB 14; Length 8;  
 Best Local Similarity 100.0%; Pred.No. 1.2e+06;  
 Matches 6; Conservative 0; Mismatches 0; Indels 0;

QY 731 KPKDEL 736  
 Db 1 KPKDEL 6

RESULT 9  
 US-10-360-836-69  
 ; Sequence 69, Application US/10360836  
 ; Publication No. US20030185854A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Zavala, Fidel  
 ; APPLICANT: Birkett, Ashley  
 ; TITLE OF INVENTION: USE OF RECOMBINANT HEPATITIS B CORE  
 ; TITLE OF INVENTION: PARTICLES TO DEVELOP VACCINES AGAINST INFECTIOUS PATHOGENS  
 ; TITLE OF INVENTION: AND MALINGNANCIES  
 ; FILE REFERENCE: 5986/1J876  
 ; CURRENT APPLICATION NUMBER: US/10/360,836  
 ; CURRENT FILING DATE: 2003-02-07  
 ; PRIOR APPLICATION NUMBER: 60/354,963  
 ; PRIOR FILING DATE: 2002-02-08  
 ; NUMBER OF SEQ ID NOS: 86  
 ; SOFTWARE: FastSeq for Windows Version 3.0  
 ; SEQ ID NO 69  
 ; LENGTH: 8  
 ; TYPE: PRT  
 ; ORGANISM: Plasmodium  
 US-10-360-836-69

Query Match 0.8%; Score 6; DB 14; Length 8;  
 Best Local Similarity 100.0%; Pred.No. 1.2e+06;  
 Matches 6; Conservative 0; Mismatches 0; Indels 0;

QY 731 KPKDEL 736  
 Db 1 KPKDEL 6

RESULT 10  
 US-10-367-674-129  
 ; Sequence 129, Application US/10367674  
 ; Publication No. US20040127884A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Rothman, James E.  
 ; APPLICANT: Hartl, F. Ulrich  
 ; APPLICANT: Hoe, Mee H.  
 ; APPLICANT: Houghton, Alan  
 ; APPLICANT: Takechi, Yoshizumi  
 ; APPLICANT: Mayhew, Mark  
 ; TITLE OF INVENTION: Heat Shock Protein-Based Vaccines and Immunotherapies  
 ; FILE REFERENCE: 11746/4610211  
 ; CURRENT APPLICATION NUMBER: US/10/367,674  
 ; CURRENT FILING DATE: 2003-02-14  
 ; PRIOR APPLICATION NUMBER: US 10/170,738  
 ; PRIOR FILING DATE: 2002-06-13  
 ; PRIOR APPLICATION NUMBER: US 09/552,868  
 ; PRIOR FILING DATE: 2000-04-20  
 ; PRIOR APPLICATION NUMBER: US 09/011,645  
 ; PRIOR FILING DATE: 1998-02-13  
 ; PRIOR APPLICATION NUMBER: PCT/US96/13363  
 ; PRIOR FILING DATE: 1996-08-16  
 ; PRIOR APPLICATION NUMBER: US 60/002,490  
 ; PRIOR FILING DATE: 1995-08-18  
 ; PRIOR APPLICATION NUMBER: US 60/002,479  
 ; PRIOR FILING DATE: 1995-08-18  
 ; NUMBER OF SEQ ID NOS: 349  
 ; SOFTWARE: WordPerfect 8.0 for Windows  
 ; SEQ ID NO 129  
 ; LENGTH: 8  
 ; TYPE: PRT  
 ; ORGANISM: Artificial Sequence  
 ; FEATURE:  
 ; OTHER INFORMATION: synthetic peptide  
 US-10-367-674-129

Query Match 0.8%; Score 6; DB 16; Length 8;  
 Best Local Similarity 100.0%; Pred.No. 1.2e+06;  
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 731 KPKDEL 736  
 Db 1 KPKDEL 6

RESULT 11  
 US-10-653-624-4  
 ; Sequence 4, Application US/10653624  
 ; Publication No. US20040131594A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: McMichael, Andrew  
 ; APPLICANT: Hill, Adrian V.S.  
 ; APPLICANT: Gilbert, Sarah C.  
 ; APPLICANT: Schneider, Joerg  
 ; APPLICANT: Plebanski, Magdalena  
 ; APPLICANT: Hanke, Tomas  
 ; APPLICANT: Smith, Geoffrey L.  
 ; APPLICANT: Blanchard, Tom  
 ; TITLE OF INVENTION: Methods and Reagents for Vaccination  
 ; FILE REFERENCE: 2907.1000-000  
 ; CURRENT APPLICATION NUMBER: US/10/653,624  
 ; CURRENT FILING DATE: 2003-09-02  
 ; PRIOR APPLICATION NUMBER: US/09/454,204A  
 ; PRIOR FILING DATE: 1999-12-09  
 ; PRIOR APPLICATION NUMBER: PCT/GB98/01681  
 ; PRIOR FILING DATE: 1998-06-09  
 ; PRIOR APPLICATION NUMBER: GB 97 11957.2  
 ; PRIOR FILING DATE: 1997-06-09  
 ; NUMBER OF SEQ ID NOS: 78



```
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 4
; LENGTH: 8
; TYPE: PRT
; ORGANISM: Unknown
; FEATURE:
; OTHER INFORMATION: CTL Epitope of the Malaria String
US-10-653-624-4

Query Match          0.8%; Score 6; DB 16; Length 8;
Best Local Similarity 100.0%; Pred. No. 1.2e+06;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 731 KPKDEL 736
Db 1 KPKDEL 6

RESULT 12
US-10-777-053-768
; Sequence 768, Application US/10777053
; Publication No. US20040132088A1
; GENERAL INFORMATION:
; APPLICANT: Simard, John J. L.
; APPLICANT: Diamond, David C.
; APPLICANT: Qiu, Zhiyong
; APPLICANT: Lei, Xiang-Dong
; TITLE OF INVENTION: EXPRESSION VECTORS ENCODING EPITOPES OF
; TITLE OF INVENTION: TARGET-ASSOCIATED ANTIGENS AND METHODS FOR THEIR DESIGN
; FILE REFERENCE: MANK:022C1
; CURRENT APPLICATION NUMBER: US/10/777,053
; CURRENT FILING DATE: 2004-02-10
; PRIOR APPLICATION NUMBER: 10/292,413
; PRIOR FILING DATE: 2002-11-07
; PRIOR APPLICATION NUMBER: 60/336,968
; PRIOR FILING DATE: 2001-11-07
; NUMBER OF SEQ ID NOS: 979
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 768
; LENGTH: 8
; TYPE: PRT
; ORGANISM: Plasmodium Falciparum
US-10-777-053-768

Query Match          0.8%; Score 6; DB 16; Length 8;
Best Local Similarity 100.0%; Pred. No. 1.2e+06;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 731 KPKDEL 736
Db 1 KPKDEL 6

RESULT 13
US-10-006-869-2953
; Sequence 2953, Application US/10006869
; Publication No. US20030082166A1
; GENERAL INFORMATION:
; APPLICANT: Blaschuk, Orest W.
; APPLICANT: Symonds, James Matthew
; APPLICANT: Gour, Barbara J.
; TITLE OF INVENTION: COMPOUNDS AND METHODS FOR MODULATING NONCLASSICAL
; TITLE OF INVENTION: CADHERIN-MEDIATED FUNCTIONS
; FILE REFERENCE: 100086.407C7
; CURRENT APPLICATION NUMBER: US/10/006,869
; CURRENT FILING DATE: 2001-12-03
; NUMBER OF SEQ ID NOS: 4052
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 2953
; LENGTH: 5
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: peptide
US-09-800-433-6

Query Match          0.7%; Score 5; DB 9; Length 6;
Best Local Similarity 100.0%; Pred. No. 1.2e+06;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 691 DDLVK 695
Db 1 DDLVK 5

RESULT 14
US-10-395-032-2953
; Sequence 2953, Application US/10395032
; Publication No. US20030229199A1
; GENERAL INFORMATION:
; APPLICANT: Blaschuk, Orest W.
; APPLICANT: Symonds, James Matthew
; APPLICANT: Gour, Barbara J.
; TITLE OF INVENTION: COMPOUNDS AND METHODS FOR MODULATING NONCLASSICAL
; TITLE OF INVENTION: CADHERIN-MEDIATED FUNCTIONS
; FILE REFERENCE: 100086.407C9
; CURRENT APPLICATION NUMBER: US/10/395,032
; CURRENT FILING DATE: 2003-03-21
; NUMBER OF SEQ ID NOS: 4052
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 2953
; LENGTH: 5
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Representative cyclic modulating agent based on
; OTHER INFORMATION: Protocadherin cell adhesion recognition sequence
US-10-395-032-2953

Query Match          0.7%; Score 5; DB 15; Length 5;
Best Local Similarity 100.0%; Pred. No. 1.2e+06;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 691 DDLVK 695
Db 1 DDLVK 5

RESULT 15
US-09-800-433-6
; Sequence 6, Application US/09800433
; Patent No. US20020106378A1
; GENERAL INFORMATION:
; APPLICANT: O'Hare and Elliott
; TITLE OF INVENTION: Materials and methods for intracellular transport and
; TITLE OF INVENTION: their uses
; FILE REFERENCE: 49408
; CURRENT APPLICATION NUMBER: US/09/800,433
; CURRENT FILING DATE: 2000-03-05
; PRIOR APPLICATION NUMBER: 09/395,344
; PRIOR FILING DATE: 1999-09-13
; NUMBER OF SEQ ID NOS: 11
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 6
; LENGTH: 6
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: peptide
US-09-800-433-6

Query Match          0.7%; Score 5; DB 9; Length 6;
Best Local Similarity 100.0%; Pred. No. 1.2e+06;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

QY 63 RSRAA 67  
|||  
Db 1 RSRAA 5

Search completed: August 17, 2004, 10:18:36  
Job time : 48 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: August 17, 2004, 10:13:20 ; Search time 20 Seconds  
(without alignments)

3539.848 Million cell updates/sec

Title: US-10-045-815-4

Perfect score: 736

Sequence: 1 MAVRALKLLTLLAVAAAS.....PPEPAQESLSGSESKPKDEL 736

Scoring table: OLIGO

Gapop 60.0 , Gapext 60.0

Searched: 283366 seqs, 96191526 residues

Word size : 0

Total number of hits satisfying chosen parameters: 605

Minimum DB seq length: 0

Maximum DB seq length: 8

Post-processing: Listing first 45 summaries

Database : PIR 78.\*

1: pir1.\*

2: pir2.\*

3: pir3.\*

4: pir4.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB ID	Description
1	5	0.7	6	2 C22565	R-phycoerythrin be
2	5	0.7	6	2 PT0605	T-cell receptor be
3	4	0.5	5	2 PT0703	T-cell receptor be
4	4	0.5	5	2 PT0701	T-cell receptor be
5	4	0.5	6	2 A11490	pyruvate kinase (E
6	4	0.5	6	2 B33932	IG mu chain D regi
7	4	0.5	6	2 PT0604	T-cell receptor be
8	4	0.5	6	2 PT0668	T-cell receptor be
9	4	0.5	7	2 JN0859	peptidyl-di-peptida
10	4	0.5	7	2 PT0654	T-cell receptor be
11	4	0.5	7	2 PT0543	T-cell receptor be
12	4	0.5	7	2 PT0689	T-cell receptor be
13	4	0.5	7	2 PT0719	T-cell receptor be
14	4	0.5	8	2 A32523	peptidyl-di-peptida
15	4	0.5	8	2 JS0315	leucokinin V - Mad
16	4	0.5	8	2 PT0725	T-cell receptor be
17	4	0.5	8	2 S68325	blood cell protein
18	3	0.4	3	3 PT0571	T-cell receptor be
19	3	0.4	4	2 A37832	phenol 2-monooxyge
20	3	0.4	4	2 S17255	ribosomal protein
21	3	0.4	4	2 JQ1273	neuropeptide Antho
22	3	0.4	4	2 E44823	synaptosomal-assoc
23	3	0.4	4	2 PT0633	T-cell receptor be
24	3	0.4	4	2 PT0711	T-cell receptor be
25	3	0.4	4	2 PT0661	T-cell receptor be
26	3	0.4	4	2 PT0698	T-cell receptor be
27	3	0.4	4	2 PT0677	T-cell receptor be
28	3	0.4	4	2 PT0706	T-cell receptor be
29	3	0.4	4	2 PT0675	T-cell receptor be

30	3	0.4	4	2 PT0721	T-cell receptor be
31	3	0.4	4	2 PT0566	protein-glutamine
32	3	0.4	4	2 A26209	pallidipin - assas
33	3	0.4	4	2 S55238	spinal cord peptid
34	3	0.4	5	2 C23751	copper resistance
35	3	0.4	5	2 A41225	copper resistance
36	3	0.4	5	2 B41225	URE2 protein - Xan
37	3	0.4	5	2 S70154	biotin B - Citroba
38	3	0.4	5	2 I40698	tram protein - Esc
39	3	0.4	5	2 A32014	alkanal monooxygen
40	3	0.4	5	2 A44955	hypothetical prote
41	3	0.4	5	2 T14910	alpha-amylase - ri
42	3	0.4	5	2 S51077	ribulose-bisphosph
43	3	0.4	5	2 PS0324	acid proteinase li
44	3	0.4	5	2 B37988	actin I - malaria
45	3	0.4	5	2 B45525	

ALIGNMENTS

RESULT 1

C22565  
R-phycoerythrin beta-1 chain - red alga (Gastrocloonium coulteri) (fragment)  
C;Species: Gastrocloonium coulteri  
C;Date: 07-Mar-1988 #sequence\_revision 07-Mar-1988 #text\_change 23-Mar-1993  
C;Accession: C22565  
R;Klotz, A.V.; Glazer, A.N.  
J. Biol. Chem. 260, 4856-4863, 1985  
A;Title: Characterization of the bilin attachment sites in R-phycoerythrin.  
A;Reference number: A22565; MUID:85182601; PMID:3886544  
A;Accession: C22565  
A;Molecule type: protein  
A;Residues: 1-6 <KLO>

Query Match 0.7%; Score 5; DB 2; Length 6;  
Best Local Similarity 100.0%; Pred. No. 2.8e+05;  
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 117 AACLR 121  
Db 2 AACLR 6

RESULT 2

PT0605  
T-cell receptor beta chain V-D-J region (120-11L) - mouse (fragment)  
C;Species: Mus musculus (house mouse)  
C;Date: 17-Jul-1992 #sequence\_revision 17-Jul-1992 #text\_change 30-May-1997  
C;Accession: PT0605  
R;Feeney, A.J.  
J. Exp. Med. 174, 115-124, 1991  
A;Title: Junctional sequences of fetal T cell receptor beta chains have few N regions.  
A;Reference number: PT0509; MUID:91277601; PMID:1711558  
A;Accession: PT0605  
A;Status: translation not shown  
A;Molecule type: mRNA  
A;Residues: 1-6 <PEE>  
A;Experimental source: newborn thymus, strain BALB/c  
C;Keywords: T-cell receptor

Query Match 0.7%; Score 5; DB 2; Length 6;  
Best Local Similarity 100.0%; Pred. No. 2.8e+05;  
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 98 ASGAG 102  
Db 1 ASGAG 5

RESULT 3

PT0703  
T-cell receptor beta chain V-D-J region (135-1F) - mouse (fragment)

C;Species: Mus musculus (house mouse)  
 C;Date: 17-Jul-1992 #sequence\_revision 17-Jul-1992 #text\_change 30-May-1997  
 C;Accession: PT0703  
 R;Feeney, A.J.  
 J. Exp. Med. 174, 115-124, 1991  
 A;Title: Junctional sequences of fetal T cell receptor beta chains have few N regions.  
 A;Reference number: PT0509; MUID:91277601; PMID:1711558  
 A;Accession: PT0703  
 A;Status: translation not shown  
 A;Molecule type: DNA  
 A;Residues: 1-5 <FEE>  
 A;Experimental source: newborn thymus, strain BALB/c  
 C;Keywords: T-cell receptor

Query Match 0.5%; Score 4; DB 2; Length 5;  
 Best Local Similarity 100.0%; Pred. No. 2.8e+05;  
 Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 496 SGD 499  
 DB 2 SGD 5

RESULT 4  
 PT0701  
 T-cell receptor beta chain V-D-J region - mouse (fragment)  
 C;Species: Mus musculus (house mouse)  
 C;Date: 17-Jul-1992 #sequence\_revision 17-Jul-1992 #text\_change 30-May-1997  
 C;Accession: PT0658; PT0701  
 R;Feeney, A.J.  
 J. Exp. Med. 174, 115-124, 1991  
 A;Title: Junctional sequences of fetal T cell receptor beta chains have few N regions.  
 A;Reference number: PT0509; MUID:91277601; PMID:1711558  
 A;Accession: PT0658  
 A;Status: translation not shown  
 A;Molecule type: mRNA  
 A;Residues: 1-5 <FEE>  
 A;Experimental source: day 4 postnatal thymus, strain BALB/c, 121-3E  
 A;Accession: PT0701  
 A;Status: translation not shown  
 A;Molecule type: DNA  
 A;Residues: 1-5 <FEE>  
 A;Experimental source: newborn thymus, strain BALB/c, 161-2E  
 C;Keywords: T-cell receptor

Query Match 0.5%; Score 4; DB 2; Length 5;  
 Best Local Similarity 100.0%; Pred. No. 2.8e+05;  
 Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 652 SSGT 655  
 DB 2 SSGT 5

RESULT 5  
 A11490  
 pyruvate kinase (EC 2.7.1.40) - pig (fragment)  
 C;Species: Sus scrofa domestica (domestic pig)  
 C;Date: 05-Jun-1987 #sequence\_revision 05-Jun-1987 #text\_change 03-Mar-1995  
 C;Accession: A11490  
 R;Hjeltnist, G.; Andersson, J.; Edlund, B.; Engstrom, L.  
 Biochem. Biophys. Res. Commun. 61, 559-563, 1974  
 A;Title: Amino acid sequence of a (32-P)phosphopeptide from pig liver pyruvate kinase ph  
 A;Reference number: A11490; MUID:75127438; PMID:4375989  
 C;Accession: A11490  
 A;Molecule type: protein  
 A;Residues: 1-6 <HUE>  
 A;Experimental source: liver  
 C;Keywords: glycolysis; phosphotransferase

Query Match 0.5%; Score 4; DB 2; Length 6;  
 Best Local Similarity 100.0%; Pred. No. 2.8e+05;  
 Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 114 LRRA 117  
 DB 1 LRRA 4

## RESULT 6

B33932  
 Ig mu chain D region (D23) - mouse  
 C;Species: Mus musculus (house mouse)  
 C;Date: 03-Mar-1990 #sequence\_revision 03-Mar-1990 #text\_change 16-Aug-1996  
 C;Accession: B33932  
 R;Baccala, R.; Vo Quang, T.; Gilbert, M.; Ternynck, T.; Avrameas, S.  
 Proc. Natl. Acad. Sci. U.S.A. 86, 4624-4628, 1989  
 A;Title: Two murine natural polyreactive autoantibodies are encoded by nonmutated germ-l  
 A;Reference number: A33932; MUID:89282823; PMID:2499887  
 C;Accession: B33932  
 A;Status: preliminary  
 A;Molecule type: mRNA  
 A;Residues: 1-6 <BAC>  
 A;Cross-references: GB:M27107  
 C;Keywords: immunoglobulin

Query Match 0.5%; Score 4; DB 2; Length 6;  
 Best Local Similarity 100.0%; Pred. No. 2.8e+05;  
 Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 71 LRRL 74  
 DB 3 LRRL 6

## RESULT 7

PT0604  
 T-cell receptor beta chain V-D-J region (111-1A) - mouse (fragment)  
 C;Species: Mus musculus (house mouse)  
 C;Date: 17-Jul-1992 #sequence\_revision 17-Jul-1992 #text\_change 30-May-1997  
 C;Accession: PT0604  
 R;Feeney, A.J.  
 J. Exp. Med. 174, 115-124, 1991  
 A;Title: Junctional sequences of fetal T cell receptor beta chains have few N regions.  
 A;Reference number: PT0509; MUID:91277601; PMID:1711558  
 A;Accession: PT0604  
 A;Status: translation not shown  
 A;Molecule type: mRNA  
 A;Residues: 1-6 <FEE>  
 A;Experimental source: newborn thymus, strain BALB/c  
 C;Keywords: T-cell receptor

Query Match 0.5%; Score 4; DB 2; Length 6;  
 Best Local Similarity 100.0%; Pred. No. 2.8e+05;  
 Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 496 SGD 499  
 DB 2 SGD 5

## RESULT 8

PT0668  
 T-cell receptor beta chain V-D-J region (121-3BB) - mouse (fragment)  
 C;Species: Mus musculus (house mouse)  
 C;Date: 17-Jul-1992 #sequence\_revision 17-Jul-1992 #text\_change 30-May-1997  
 C;Accession: PT0668  
 R;Feeney, A.J.  
 J. Exp. Med. 174, 115-124, 1991  
 A;Title: Junctional sequences of fetal T cell receptor beta chains have few N regions.  
 A;Reference number: PT0509; MUID:91277601; PMID:1711558  
 A;Accession: PT0668  
 A;Status: translation not shown  
 A;Molecule type: mRNA  
 A;Residues: 1-6 <FEE>  
 A;Experimental source: day 4 postnatal thymus, strain BALB/c

C;Keywords: T-cell receptor

Query Match 0.5%; Score 4; DB 2; Length 6;  
Best Local Similarity 100.0%; Pred. No. 2.8e+05;  
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 652 SSGT 655  
|||  
Db 2 SSGT 5

#### RESULT 9

JN0859  
peptidyl-dipeptidase A inhibitory peptide C105 - striped bonito  
C;Species: Sarda orientalis (striped bonito)  
C;Date: 10-Mar-1994 #sequence\_revision 10-Mar-1994 #text\_change 07-May-1999  
C;Accession: JN0859  
R;Matsumura, N.; Fujii, M.; Takeda, Y.; Shimizu, T.  
BioSci. Biotechnol. Biochem. 57, 1743-1744, 1993  
A;Title: Isolation and characterization of angiotensin I-converting enzyme inhibitory peptide  
A;Reference number: JN0859; PMID:94080036; PMID:7764272  
A;Accession: JN0859  
A;Molecule type: protein  
A;Residues: 1-7 <MA>  
A;Experimental source: intestine  
C;Comment: The carboxyl-terminus is essential for the protein's expression of angiotensin  
C;Superfamily: bradykinin-potentiating peptide  
C;Keywords: angiotensin-converting enzyme inhibitor

Query Match 0.5%; Score 4; DB 2; Length 7;  
Best Local Similarity 100.0%; Pred. No. 2.8e+05;  
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 158 KLEK 161  
|||  
Db 4 KLEK 7

#### RESULT 10

PT0654  
T-cell receptor beta chain V-D-J region (121-1BK) - mouse (fragment)  
C;Species: Mus musculus (house mouse)  
C;Date: 17-Jul-1992 #sequence\_revision 17-Jul-1992 #text\_change 30-May-1997  
C;Accession: PT0654  
R;Feeney, A.J.  
J. Exp. Med. 174, 115-124, 1991  
A;Title: Junctional sequences of fetal T cell receptor beta chains have few N regions.  
A;Reference number: PT0509; PMID:91277601; PMID:1711558  
A;Accession: PT0654  
A;Status: translation not shown  
A;Molecule type: mRNA  
A;Residues: 1-7 <FEE>  
A;Experimental source: day 4 postnatal thymus, strain BALB/c  
C;Keywords: T-cell receptor

Query Match 0.5%; Score 4; DB 2; Length 7;  
Best Local Similarity 100.0%; Pred. No. 2.8e+05;  
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 98 ASGA 101  
|||  
Db 1 ASGA 4

#### RESULT 11

PT0543  
T-cell receptor beta chain V-D-J region (126-1BE) - mouse (fragment)  
C;Species: Mus musculus (house mouse)  
C;Date: 17-Jul-1992 #sequence\_revision 17-Jul-1992 #text\_change 30-May-1997  
C;Accession: PT0543  
R;Feeney, A.J.  
J. Exp. Med. 174, 115-124, 1991  
A;Title: Junctional sequences of fetal T cell receptor beta chains have few N regions.

A;Reference number: PT0509; PMID:91277601; PMID:1711558

A;Accession: PT0543

A;Status: translation not shown

A;Molecule type: mRNA

A;Residues: 1-7 <FEE>

A;Experimental source: day 18 fetal thymus, strain BALB/c

C;Keywords: T-cell receptor

Query Match 0.5%; Score 4; DB 2; Length 7;  
Best Local Similarity 100.0%; Pred. No. 2.8e+05;  
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 496 SGDQ 499  
|||  
Db 2 SGDQ 5

#### RESULT 12

PT0689  
T-cell receptor beta chain V-D-J region (140-1AF) - mouse (fragment)  
C;Species: Mus musculus (house mouse)  
C;Date: 17-Jul-1992 #sequence\_revision 17-Jul-1992 #text\_change 30-May-1997  
C;Accession: PT0689  
R;Feeney, A.J.  
J. Exp. Med. 174, 115-124, 1991  
A;Title: Junctional sequences of fetal T cell receptor beta chains have few N regions.  
A;Reference number: PT0509; PMID:91277601; PMID:1711558  
A;Accession: PT0689  
A;Status: translation not shown  
A;Molecule type: DNA  
A;Residues: 1-7 <FEE>  
A;Experimental source: day 18 fetal thymus, strain BALB/c  
C;Keywords: T-cell receptor

Query Match 0.5%; Score 4; DB 2; Length 7;  
Best Local Similarity 100.0%; Pred. No. 2.8e+05;  
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 496 SGDQ 499

Db 2 SGDQ 5

|||

|||

|||

|||

|||

|||

|||

|||

|||

|||

|||

|||

|||

|||

|||

|||

|||

|||

|||

|||

|||

|||

|||

|||

|||

|||

|||

|||

|||

|||

|||

|||

|||

|||

|||

|||

|||

|||

|||

|||

|||

|||

|||

|||

|||

|||

|||

|||

|||

|||

|||

## RESULT 14

A32523  
 peptidyl-dipeptidase A (EC 3.4.15.1) - bovine (fragment)  
 N:Alternate names: angiotensin I-converting enzyme; peptidyl-dipeptidase I  
 C:Species: Bos primigenius taurus (cattle)  
 C:Date: 18-Oct-1989 #sequence\_revision 18-Oct-1989 #text\_change 11-May-2000  
 C:Accession: A32523  
 R:Harris, R.B. Biol. 198, 513-521, 1986  
 Adv. Exp. Med. Biol. 198, 513-521, 1986  
 A:Title: Isolation and sequencing of an active-site peptide from angiotensin I-converting  
 A:Reference number: A32523; PMID:87123961; PMID:3028071  
 A:Accession: A32523  
 A:Molecule type: protein  
 A:Residues: 1-8 <HAR>  
 C:Superfamily: mammalian peptidyl-dipeptidase A  
 C:Keywords: alternative splicing; blood pressure control; peptidyl-dipeptide hydrolase; z

Query Match 0.5%; Score 4; DB 2; Length 8;  
 Best Local Similarity 100.0%; Pred. No. 2.8e+05;  
 Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 284 TELA 287

||||

Db 2 TELA 5

## RESULT 15

JS0315  
 leucokinin V - Madeira cockroach  
 C:Species: Leucophaea maderae (Madeira cockroach)  
 C:Date: 07-Sep-1990 #sequence\_revision 07-Sep-1990 #text\_change 20-Jun-2000  
 C:Accession: JS0315  
 R:Holman, G.M.; Cook, B.J.; Nachman, R.J.  
 Comp. Biochem. Physiol. C 88, 27-30, 1987  
 A:Title: Isolation, primary structure, and synthesis of leucokinin V and VI: myotropic  
 A:Reference number: JS0315  
 A:Accession: JS0315  
 A:Molecule type: protein  
 A:Residues: 1-8 <HOL>  
 C:Comment: Leucokinin, a family of cephalomyotropic peptides, stimulate contractile act  
 C:Keywords: amidated carboxyl end; cephalomyotropic peptide  
 F,8/Modified site: amidated carboxyl end (Gly) #status experimental

Query Match 0.5%; Score 4; DB 2; Length 8;  
 Best Local Similarity 100.0%; Pred. No. 2.8e+05;  
 Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 650 GFSS 653

||||

Db 3 GFSS 6

Search completed: August 17, 2004, 10:17:08  
 Job time : 21 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: August 17, 2004, 10:09:40 ; Search time 13 Seconds  
(without alignments)

2947.972 Million cell updates/sec

Title: US-10-045-815-4

Perfect score: 736

Sequence: 1 MAVRALKLLTLLAVAAAS.....PPEPAQESLSGSESKPKDEL 736

Scoring table: OLIGO

Gapop 60.0 , Gapext 60.0

Searched: 141681 seqs, 52070155 residues

Word size : 0

Total number of hits satisfying chosen parameters: 158

Minimum DB seq length: 0

Maximum DB seq length: 8

Post-processing: listing first 45 summaries

Database : SwissProt\_42:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Query Match	Score	Length	ID	Description
1	5	0.7	8	1 FAR4 MACRS	P83277 macrobrachi
2	4	0.5	5	1 TPIS CANFA	P54714 canis famil
3	4	0.5	8	1 LCK2 LEUMA	P21141 leucophaea
4	4	0.5	8	1 LCK5 LEUMA	P19987 leucophaea
5	3	0.4	4	1 E0S1 HUMAN	P02731 homo sapien
6	3	0.4	4	1 FKA ANTEL	P38705 anthopleura
7	3	0.4	4	1 RW01 YEAST	P36515 saccharomyc
8	3	0.4	4	1 B10B CITR	P12997 citrobacter
9	3	0.4	5	1 PAP2 FARMA	P81864 pardachirus
10	3	0.4	5	1 PSK DAUCA	P58261 gaurus caro
11	3	0.4	5	1 RE11 LITRU	P82070 litoria rub
12	3	0.4	5	1 RE21 LITRU	P82071 litoria rub
13	3	0.4	5	1 RE32 LITRU	P82073 litoria rub
14	3	0.4	5	1 SUGA ACHDO	P19991 acheta dome
15	3	0.4	5	1 TRM3 ECOLI	P13973 escherichia
16	3	0.4	5	1 UXA4 CHLTR	P38005 chlamydia t
17	3	0.4	6	1 ACPH RABIT	P25154 oryctolagus
18	3	0.4	6	1 ASP2 LACSN	P82655 lactobacill
19	3	0.4	6	1 FARP MONEX	P41966 moniezia ex
20	3	0.4	6	1 LOX1 LOCOM	P41491 locusta mig
21	3	0.4	6	1 UN06 CLOPA	P81351 clostridium
22	3	0.4	7	1 ALL2 CARMA	P81805 carcinus ma
23	3	0.4	7	1 BRHP CONIM	P58803 conus imper
24	3	0.4	7	1 CARP MYTED	P10420 mytilus edu
25	3	0.4	7	1 CCF1 ENTEA	P20104 enterococcu
26	3	0.4	7	1 CIA ENTEA	P11932 enterococcu
27	3	0.4	7	1 GRFP MOUSE	P59025 mus musculu
28	3	0.4	7	1 LANC CARUI	P36960 carnobacter
29	3	0.4	7	1 MNF1 LEPEDE	P42984 leptinotars
30	3	0.4	7	1 PP2 LYCES	P83379 lycopersico
31	3	0.4	7	1 TPFY PACDA	P83455 pachymedusa
32	3	0.4	7	1 UF03 MOUSE	P38641 mus musculu
33	3	0.4	7	1 UH11 RAT	P56576 rattus norv

```

34 3 0.4 7 1 UN06 PINPS P81675 pinus pinas
35 3 0.4 8 1 AKH MELML P25423 melolontha
36 3 0.4 8 1 ALL5 CARMA P81818 carcinus ma
37 3 0.4 8 1 ALL6 CARMA P81819 carcinus ma
38 3 0.4 8 1 ALL1 CYDPO P82152 cydia pomon
39 3 0.4 8 1 ALL5 CALVO P41841 calliphora
40 3 0.4 8 1 ALL5 CYDPO P82156 cydia pomon
41 3 0.4 8 1 ALL6 CYDPO P82157 cydia pomon
42 3 0.4 8 1 ALL7 CARMA P81809 carcinus ma
43 3 0.4 8 1 ALL8 CARMA P81811 carcinus ma
44 3 0.4 8 1 ALL9 CARMA P81812 carcinus ma
45 3 0.4 8 1 ANG2 BOTUA Q10582 bothrops ja

```

#### ALIGNMENTS

```

RESULT 1
FAR4 MACRS
ID FAR4 MACRS STANDARD; PRT; 8 AA.
AC P83277; 2003 (Rel. 41, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE FMRamide-like neuropeptide FLPA (APALRLRF-amide).
OS Macrobrachium rosenbergii (Giant fresh water prawn).
OC Eukaryota; Metazoa; Arthropoda; Crustacea; Malacostraca;
OC Eumalacostraca; Eucarida; Decapoda; Pleocyemata; Caridea;
OC Palaemonoidea; Palaemonidae; Macrobrachium.
OX NCBI_TaxID=79674;
RN [1]
RP SEQUENCE, AND MASS SPECTROMETRY.
RC TISSUE=Eyestalk;
RA Sithigonggul W.;
RA Sithigonggul P.; Sarathongkum W., Jaidechoey S., Longyant S.,
RT "Novel FMRamide-like neuropeptides from the eyestalk of the giant
RT freshwater prawn Macrobrachium rosenbergii.";
RL Comp. Biochem. Physiol. 120B:587-595(1998).
CC -1- MASS SPECTROMETRY; MW=943; METHOD=WALDI.
CC -1- SIMILARITY: Belongs to the FARP (FMRamide related peptide)
CC family.
DR GO:0007218; P:neuropeptide signaling pathway; TAS.
KW Neuropeptide; Amidation.
FT MOD RES 8 8 AMIDATION.
SQ SEQUENCE 8 AA; 943 MW; 9CD40734072DC76D CRC64;

```

Query Match 0.7%; Score 5; DB 1; Length 8;  
Best Local Similarity 100.0%; Pred. No. 1.4e+05;  
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

Qy 70 ALRLR 74
    |||||
Db 3 ALRLR 7

```

```

RESULT 2
TPIS CANFA
ID TPIS CANFA STANDARD; PRT; 5 AA.
AC P54714;
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 15-JUL-1998 (Rel. 36, Last annotation update)
DE Triosephosphate isomerase (EC 5.3.1.1) (TIM) (Fragment).
GN TP11.
OS Canis familiaris (Dog).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Carnivora; Fissipedia; Canidae; Canis.
OX NCBI_TaxID=9615;
RN [1]
RP SEQUENCE.
RC TISSUE=Heart;
RX MEDLINE=98163340; PubMed=9504812;
RA Dunn M.J., Corbett J.M., Wheeler C.H.;

```



RT "HSC-2DPAGE and the two-dimensional gel electrophoresis database of  
 RT dog heart proteins.";  
 RL Electrophoresis 18:2795-2802(1997).  
 CC -!- CATALYTIC ACTIVITY: D-glyceraldehyde 3-phosphate = glycerone  
 CC phosphate.  
 CC -!- PATHWAY: Plays an important role in several metabolic pathways.  
 CC -!- SUBUNIT: Homodimer (By similarity).  
 CC -!- SIMILARITY: Belongs to the triosephosphate isomerase family.  
 DR HSC-2DPAGE; P54714; DOG.  
 DR InterPro; IPR000652; Triophos.ismrse.  
 DR PROSITE; PS00171; TIM; PARTIAL.  
 KW Isomerase; Glycolysis; Gluconogenesis; Fatty acid biosynthesis;  
 KW Pentose shunt.  
 FT NON\_TER 1 1  
 FT NON\_TER 5 5  
 SQ SEQUENCE 5 AA; 550 MW; 64444862C9A00000 CRC64;

Query Match 0.5%; Score 4; DB 1; Length 5;  
 Best Local Similarity 100.0%; Pred. No. 1.4e+05;  
 Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 170 FVGN 173  
 Db ||||  
 1 FVGN 4

RESULT 3  
 LCK2\_LEUMA  
 ID LCK2\_LEUMA STANDARD; PRT; 8 AA.  
 AC P21141;  
 DT 01-MAY-1991 (Rel. 18, Created)  
 DT 01-MAY-1991 (Rel. 18, Last sequence update)  
 DT 10-OCT-2003 (Rel. 42, Last annotation update)  
 DE Leucokinin II (L-II).  
 OS Leucophaea maderae (Madeira cockroach).  
 OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;  
 OC Neoptera; Orthopteroidea; Dictyoptera; Blattaria; Blaberoidea;  
 OC Blaberoidea; Leucophaea.  
 OX NCBI\_TaxID=6988;  
 RN [1]  
 RP SEQUENCE, AND SYNTHESIS.  
 RT TISSUE=Head;  
 RA Holman G.M., Cook B.J., Nachman R.J.;  
 RT "Isolation, primary structure and synthesis of two neuropeptides  
 RT from Leucophaea maderae: members of a new family of  
 RT Cephalomyotropins.";  
 RL Comp. Biochem. Physiol. 84C:205-211(1986).  
 CC -!- FUNCTION: This cephalomyotropic peptide stimulates contractile  
 CC activity of cockroach proctodeum (hindgut).  
 CC -!- SUBCELLULAR LOCATION: Secreted.  
 KW Neuropeptide; Amidation.  
 FT MOD\_RES 8 8  
 FT MOD\_RES 8 8  
 SQ SEQUENCE 8 AA; 852 MW; DC6365A5B9C8676A CRC64;

Query Match 0.5%; Score 4; DB 1; Length 8;  
 Best Local Similarity 100.0%; Pred. No. 1.4e+05;  
 Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 650 GFSS 653  
 Db ||||  
 3 GFSS 6

RESULT 4  
 LCK5\_LEUMA  
 ID LCK5\_LEUMA STANDARD; PRT; 8 AA.  
 AC P19987;  
 DT 01-FEB-1991 (Rel. 17, Created)  
 DT 01-FEB-1991 (Rel. 17, Last sequence update)  
 DT 10-OCT-2003 (Rel. 42, Last annotation update)  
 DE Leucokinin V (L-V).  
 OS Leucophaea maderae (Madeira cockroach).  
 OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;

OC Neoptera; Orthopteroidea; Dictyoptera; Blattaria; Blaberoidea;  
 OC Blaberoidea; Leucophaea.  
 OX NCBI\_TaxID=6988;  
 RN [1]  
 RP SEQUENCE.  
 RC TISSUE=Head;  
 RX MEDLINE=87052651; PubMed=2877794;  
 RA Holman G.M., Cook B.J., Nachman R.J.;  
 RT "Isolation, primary structure, and synthesis of leucokinin V and VI:  
 RT myotropic peptides of Leucophaea maderae.";  
 RL Comp. Biochem. Physiol. 88C:27-30(1987).  
 CC -!- FUNCTION: This cephalomyotropic peptide stimulates contractile  
 CC activity of cockroach proctodeum (hindgut).  
 CC -!- SUBCELLULAR LOCATION: Secreted.  
 CC PIR; JS0315; JS0315.  
 DR Neuropeptide; Amidation.  
 KW MOD\_RES 8 8  
 FT MOD\_RES 8 8  
 SQ SEQUENCE 8 AA; 784 MW; 736365A5B9C865B8 CRC64;

Query Match 0.5%; Score 4; DB 1; Length 8;  
 Best Local Similarity 100.0%; Pred. No. 1.4e+05;  
 Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 650 GFSS 653  
 Db ||||  
 3 GFSS 6

RESULT 5  
 EOSI\_HUMAN  
 ID EOSI\_HUMAN STANDARD; PRT; 4 AA.  
 AC P02731;  
 DT 21-JUL-1986 (Rel. 01, Created)  
 DT 21-JUL-1986 (Rel. 01, Last sequence update)  
 DT 21-JUL-1986 (Rel. 01, Last annotation update)  
 DE Eosinophilotactic peptides.  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Euthera; Primates; Catarrhini; Hominidae; Homo.  
 OX NCBI\_TaxID=9606;  
 RN [1]  
 RP SEQUENCE.  
 RX MEDLINE=76078412; PubMed=1060093;  
 RA Goetzi E.J., Austen K.F.;  
 RT "Purification and synthesis of eosinophilotactic tetrapeptides of  
 RT human lung tissue: identification as eosinophil chemotactic factor of  
 RT anaphylaxis.";  
 RL Proc. Natl. Acad. Sci. U.S.A. 72:4123-4127(1975).  
 CC -!- MISCELLANEOUS: These peptides are released from mast cells in lung  
 CC (and other tissues) during hypersensitivity reactions  
 CC (anaphylaxis). Their activities, preferentially affecting  
 CC eosinophils, include chemotaxis, chemotactic deactivation, release  
 CC of enzymes, and stimulation of the hexose monophosphate shunt.  
 DR GO; GO:0006935; P:chemotaxis; IDA.  
 DR GO; GO:0006955; P:immune response; IDA.  
 DR VARIANT 1 1  
 FT V -> A (IN OTHER PEPTIDE).  
 FT /FTID=VAR 005201.  
 SQ SEQUENCE 4 AA; 390 MW; 6B05B862A0000000 CRC64;

Query Match 0.4%; Score 3; DB 1; Length 4;  
 Best Local Similarity 100.0%; Pred. No. 1.4e+05;  
 Matches 3; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 727 GSE 729  
 Db ||||  
 2 GSE 4

RESULT 6  
 FFKA\_ANTEL  
 ID FFKA\_ANTEL STANDARD; PRT; 4 AA.  
 AC P58705;  
 DT 28-FEB-2003 (Rel. 41, Created)

DT 28-FEB-2003 (Rel. 41, Last sequence update)  
 DE Antho-KAamide.  
 OS Anthopleura elegantissima (Sea anemone).  
 OC Eukaryota; Metazoa; Cnidaria; Anthozoa; Zoantharia; Actiniaria;  
 OC Nynanthaeae; Actiniidae; Anthopleura.  
 CX NCBI\_TaxID=6110;  
 RN [1]  
 RP SEQUENCE  
 RX MEDLINE=92028852; PubMed=1681803;  
 RA Nothacker H.-P., Rinehart K.L. Jr., Grimmelikhuijzen C.J.P.;  
 RT "Isolation of L-3-phenyllactyl-Phe-Lys-Ala-NH2 (Antho-KAamide), a  
 novel neuropeptide from sea anemones";  
 RL Biochem. Biophys. Res. Commun. 179:1205-1211(1991).  
 RN [2]  
 RP FUNCTION  
 RX MEDLINE=93391436; PubMed=8397415;  
 RA McFarlane I.D., Hudman D., Nothacker H.-P., Grimmelikhuijzen C.J.P.;  
 RT "The expansion behaviour of sea anemones may be coordinated by two  
 inhibitory neuropeptides, Antho-KAamide and Antho-Ramide";  
 RL Proc. R. Soc. Lond., B, Biol. Sci. 253:183-188(1993).  
 CC -/- FUNCTION: Inhibits spontaneous contractions in several muscle  
 groups. May be involved in the expansion phase of feeding  
 behaviour in sea anemones.  
 CC -/- SUBCELLULAR LOCATION: Secreted.  
 CC -/- TISSUE SPECIFICITY: Neuron specific.  
 DR PIR; JQ1273; JQ1273.  
 KW Neuropeptide; Amidation.  
 FT MOD RES 1 1 L-3-PHENYLLACTYL.  
 FT MOD RES 4 4 AMIDATION.  
 SQ SEQUENCE 4 AA; 512 MW; 6DD339CA00000000 CRC64;  
 Query Match 0.4%; Score 3; DB 1; Length 4;  
 Best Local Similarity 100.0%; Pred. No. 1.4e+05;  
 Matches 3; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 519 FKA 521  
 DB 2 FKA 4  
 RESULT 7  
 RM01\_YEAST STANDARD; PRT; 4 AA.  
 AC P36515;  
 DT 01-JUN-1994 (Rel. 29, Created)  
 DE 01-JUN-1994 (Rel. 29, Last sequence update)  
 DT 01-JUN-1994 (Rel. 29, Last annotation update)  
 DE Mitochondrial 60S ribosomal protein L1 (YmL1) (Fragment).  
 GN MRPL1.  
 OS Saccharomyces cerevisiae (Baker's yeast).  
 CC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;  
 CC Saccharomycetales; Saccharomycetaceae; Saccharomycetes.  
 CX NCBI\_TaxID=4932;  
 RN [1]  
 RP SEQUENCE.  
 RX MEDLINE=91285106; PubMed=2060626;  
 RA Grobmann L., Graack H.-R., Kruff V., Choli T., Goldschmidt-Reisin S.,  
 RA Kitakawa M.;  
 RT "Extended N-terminal sequencing of proteins of the large ribosomal  
 subunit from yeast mitochondria";  
 RL FEBS Lett. 284:51-56(1991).  
 DR PIR; S17255; S17255.  
 DR SGD; L0002681; MRPL1.  
 KW Ribosomal protein; Mitochondrion.  
 FT NON\_TER 4 4  
 SQ SEQUENCE 4 AA; 402 MW; 7771B2D5D000000000 CRC64;  
 Query Match 0.4%; Score 3; DB 1; Length 4;  
 Best Local Similarity 100.0%; Pred. No. 1.4e+05;  
 Matches 3; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 33 VTP 35

Db 2 VTP 4  
 RESULT 8  
 BIOB\_CITFR STANDARD; PRT; 5 AA.  
 ID BIOB\_CITFR  
 AC P12997;  
 DT 01-JAN-1990 (Rel. 13, Created)  
 DT 01-JAN-1990 (Rel. 13, Last sequence update)  
 DT 28-FEB-2003 (Rel. 41, Last annotation update)  
 DE Biotin synthase (EC 2.8.1.6) (Biotin synthetase) (Fragment).  
 GN BIOB.  
 OS Citrobacter freundii.  
 CC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;  
 CC Enterobacteriaceae; Citrobacter.  
 CX NCBI\_TaxID=546;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=89006280; PubMed=2371595;  
 RA Shivan D., Campbell A.;  
 RT "Transcriptional regulation and gene arrangement of Escherichia coli,  
 Citrobacter freundii and Salmonella typhimurium biotin operons";  
 RL Gene 67:203-211(1988).  
 CC -/- CATALYTIC ACTIVITY: Dethiobiotin + sulfur = biotin.  
 CC -/- PATHWAY: Biotin biosynthesis; last step.  
 CC -/- SIMILARITY: Belongs to the biotin and lipoic acid synthetases  
 family.  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 the European Bioinformatics Institute. There are no restrictions on its  
 use by non-profit institutions as long as its content is in no way  
 modified and this statement is not removed. Usage by and for commercial  
 entities requires a license agreement (See [http://www.isb-sib.ch/](http://www.isb-sib.ch/announce/)  
 or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 DR EMBL; M21922; -; NOT\_ANNOTATED\_CDS.  
 DR PIR; I40698; I40698.  
 KW Biotin biosynthesis; Iron-sulfur; Transferase.  
 FT NON\_TER 5 5  
 SQ SEQUENCE 5 AA; 532 MW; 75A5B1EDD6F00000 CRC64;  
 Query Match 0.4%; Score 3; DB 1; Length 5;  
 Best Local Similarity 100.0%; Pred. No. 1.4e+05;  
 Matches 3; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 129 AHS 131  
 DB 2 AHS 4  
 RESULT 9  
 PAP2\_PARMA STANDARD; PRT; 5 AA.  
 ID PAP2\_PARMA  
 AC P81864;  
 DT 30-MAY-2000 (Rel. 39, Created)  
 DT 30-MAY-2000 (Rel. 39, Last sequence update)  
 DT 28-FEB-2003 (Rel. 41, Last annotation update)  
 DE Pardaxin II (PXII) (Fragment).  
 OS Pardachirus marmoratus (Red sea moose sole).  
 CC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 CC Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;  
 CC Acanthomorpha; Acanthopterygii; Percomorpha; Pleuronectiformes;  
 CC Soleiidae; Soleidae; Pardachirus.  
 CX NCBI\_TaxID=31087;  
 RN [1]  
 RP SEQUENCE.  
 RC TISSUE=Skin secretion;  
 RX MEDLINE=87057369; PubMed=3782138;  
 RA Lazarovici P., Primor N., Loew L.M.;  
 RT "Purification and pore-forming activity of two hydrophobic  
 polypeptides from the secretion of the Red sea moose sole (Pardachirus

```

RT marmoratus).";
RL J. Biol. Chem. 261:16704-16713(1986).
CC -!- FUNCTION: Exhibits unusual shark repellent and surfactant
CC properties. Forms voltage-dependent, ion-permeable channels
CC in membranes. At high concentration causes cell membrane lysis.
CC -!- SUBUNIT: Monomer. In aqueous solution exists as a tetramer.
CC -!- SUBCELLULAR LOCATION: Secreted.
CC -!- SIMILARITY: Belongs to the pardaxin family.
KW Toxin.
FT NON_TER
SQ SEQUENCE 5 AA; 614 MW; 7769C9C9C8100000 CRC64;

Query Match 0.4%; Score 3; DB 1; Length 5;
Best Local Similarity 100.0%; Pred. No. 1.4e+05;
Matches 3; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 330 PFP 332
DB |||
3 PFP 5

RESULT 10
PSK_DAUCA STANDARD; PRT; 5 AA.
AC P58261,
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Phytosulfokine-alpha (PSK-alpha) [Contains: Phytosulfokine-beta (PSK-beta)].
OS Daucus carota (Carrot).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; asterids;
OC campanulids; Apiales; Apiaceae; Apioideae; Scandiceae; Daucinae;
OC Daucus.
OX NCBI_TaxID=4039;
RN [1]
RP SEQUENCE, AND IDENTIFICATION BY MASS SPECTROMETRY.
RC STRAIN=cv. US-Harumakigosun;
RX MEDLINE=20212743; PubMed=10750705;
RA Hanai H., Matsuno T., Yamamoto M., Matsubayashi Y., Kobayashi T.,
RA Kamada H., Sakagami Y.
RT "A secreted peptide growth factor, phytosulfokine, acting as a
RT stimulatory factor of carrot somatic embryo formation.";
RL Plant Cell Physiol. 41:27-32(2000).
CC -!- FUNCTION: In presence of 2,4-D, stimulates proliferation of the
CC cells, but does not stimulate differentiation into the somatic
CC embryos.
CC -!- SUBCELLULAR LOCATION: Secreted.
CC -!- PTM: Sulfation is important for activity and for the binding to a
CC putative membrane receptor (By similarity).
CC -!- SIMILARITY: Belongs to the phytosulfokine family.
KW Growth factor; Sulfation.
FT PEPTIDE 1 4 PHYTOSULFOKINE-BETA.
FT MOD_RES 1 1 SULFATION.
FT MOD_RES 3 3 SULFATION.
SQ SEQUENCE 5 AA; 687 MW; 76C1BB504B300000 CRC64;

Query Match 0.4%; Score 3; DB 1; Length 5;
Best Local Similarity 100.0%; Pred. No. 1.4e+05;
Matches 3; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 317 YTO 319
DB |||
3 YTO 5

RESULT 11
RE11_LITRU STANDARD; PRT; 5 AA.
AC P82070;
DT 28-FEB-2003 (Rel. 41, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)

```

```

DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Rubellidin 1.1.
OS Litoria rubella (Desert tree frog).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Amphibia; Batrachia; Anura; Neobatrachia; Hyloidea; Hylidae;
OC Pelodyadinae; Litoria.
OX NCBI_TaxID=104895;
RN [1]
RP SEQUENCE, AND MASS SPECTROMETRY.
RC TISSUE=Skin secretion;
RA Steinborner S.T., Wabnitz P.A., Waugh R.J., Bowie J.H., Gao C.,
RA Tyler M.J., Wallace J.C.;
RT "The structure of new peptides from the Australian red tree frog
RT 'Litoria rubella'. The skin peptide profile as a probe for the study
RT of evolutionary trends of amphibians.";
RL Aust. J. Chem. 49:955-963(1996).
CC -!- FUNCTION: Shows neither neuropeptide activity nor antibiotic
CC activity.
CC -!- SUBCELLULAR LOCATION: Secreted.
CC -!- TISSUE SPECIFICITY: Expressed by the skin dorsal glands.
CC -!- MASS SPECTROMETRY: MW=598; METHOD=FAB.
KW Amphibian defense peptide.
SQ SEQUENCE 5 AA; 598 MW; 6DD9C9CAB2A00000 CRC64;

Query Match 0.4%; Score 3; DB 1; Length 5;
Best Local Similarity 100.0%; Pred. No. 1.4e+05;
Matches 3; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 377 FFA 379
DB |||
3 FFA 5

RESULT 12
RE21_LITRU STANDARD; PRT; 5 AA.
AC P82071,
DT 28-FEB-2003 (Rel. 41, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Rubellidin 2.1.
OS Litoria rubella (Desert tree frog).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Amphibia; Batrachia; Anura; Neobatrachia; Hyloidea; Hylidae;
OC Pelodyadinae; Litoria.
OX NCBI_TaxID=104895;
RN [1]
RP SEQUENCE, AND MASS SPECTROMETRY.
RC TISSUE=Skin secretion;
RA Steinborner S.T., Wabnitz P.A., Waugh R.J., Bowie J.H., Gao C.,
RA Tyler M.J., Wallace J.C.;
RT "The structure of new peptides from the Australian red tree frog
RT 'Litoria rubella'. The skin peptide profile as a probe for the study
RT of evolutionary trends of amphibians.";
RL Aust. J. Chem. 49:955-963(1996).
CC -!- FUNCTION: Shows neither neuropeptide activity nor antibiotic
CC activity.
CC -!- SUBCELLULAR LOCATION: Secreted.
CC -!- TISSUE SPECIFICITY: Expressed by the skin dorsal glands.
CC -!- MASS SPECTROMETRY: MW=626; METHOD=FAB.
KW Amphibian defense peptide.
SQ SEQUENCE 5 AA; 626 MW; 6DD9C9CB10300000 CRC64;

Query Match 0.4%; Score 3; DB 1; Length 5;
Best Local Similarity 100.0%; Pred. No. 1.4e+05;
Matches 3; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 377 PFA 379
DB |||
3 PFA 5

RESULT 13

```

```

RE32_LITRU          STANDARD;          PRT;          5 AA.
ID RE32_LITRU
AC P82073;
DT 28-FEB-2003 (Rel. 41, Created)
DE TRAM protein (Fragment).
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DE TRAM
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Rubellidin 3.2.
OS Litoria rubella (Desert tree frog).
CC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
CC Amphibia; Batrachia; Anura; Neobatrachia; Hylidae; Hylinae;
CC Pelodyadinae; Litoria.
OX NCBI_TaxID=104895;
RN [1]
RP SEQUENCE
RC TISSUE=Skin secretion;
RA Wabnitz P.A., Bowie J.H., Tyler M.J., Wallace J.C.;
RT "Peptides from the skin glands of the Australian buzging tree frog
RT Litoria electrica. Comparison with the skin peptides from Litoria
RT rubella."
RL Aust. J. Chem. 52:639-645(1999).
CC -!- FUNCTION: Shows neither neuropeptide activity nor antibiotic
CC activity.
CC -!- SUBCELLULAR LOCATION: Secreted.
CC -!- TISSUE SPECIFICITY: Expressed by the skin dorsal glands.
KW Amphibian defense peptide.
SQ SEQUENCE 5 AA; 570 MW; 71A9C9C862A00000 CRC64;

Query Match          0.4%; Score 3; DB 1; Length 5;
Best Local Similarity 100.0%; Pred. No. 1.4e+05;
Matches 3; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 649 VGF 651
DB 1 VGF 3

RESULT 14
SUGA_ACHDO
ID SUGA_ACHDO          STANDARD;          PRT;          5 AA.
AC F19991;
DT 01-FEB-1991 (Rel. 17, Created)
DT 01-FEB-1991 (Rel. 17, Last sequence update)
DE Subesophageal ganglion pentapeptide.
OS Acheta domestica (House cricket).
CC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Psterygota;
CC Neoptera; Orthopteroidea; Orthoptera; Ensifera; Gryllidae; Gryllinae;
CC Acheta.
OX NCBI_TaxID=6937;
RN [1]
RP SEQUENCE.
RA Wicker C., Wicker C.;
RT "Isolation and structure of a peptide isolated from the
RT subesophageal ganglion of Acheta domestica (Orthoptera).";
RL Comp. Biochem. Physiol. 88C:185-187(1987).
CC -!- SUBCELLULAR LOCATION: Main peptide from the subesophageal
CC ganglia.
DR PIR; JS0319; JS0319.
SQ SEQUENCE 5 AA; 476 MW; 69D76DDDB00000 CRC64;

Query Match          0.4%; Score 3; DB 1; Length 5;
Best Local Similarity 100.0%; Pred. No. 1.4e+05;
Matches 3; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 17 AAA 19
DB 1 AAA 3

RESULT 15
TRM3_ECOLI
ID TRM3_ECOLI          STANDARD;          PRT;          5 AA.
AC F13973;

```

```

DT 01-JAN-1990 (Rel. 13, Created)
DT 01-JAN-1990 (Rel. 13, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE TRAM protein (Fragment).
GN TRAM
OS Escherichia coli.
OC Plasmid IncFII R100.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Escherichia.
OX NCBI_TaxID=562;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=88227859; PubMed=2836369;
RA Inamoto S., Yoshioka Y., Ohtsubo E.;
RT "Identification and characterization of the products from the traJ
RT and traY genes of plasmid R100."
RL J. Bacteriol. 170:2749-2757(1988).
CC -!- FUNCTION: TRANSFER GENE PROTEIN. IS INVOLVED IN THE CONJUGATION
CC PROCESS OF BACTERIAL CELLS FOR THE EXCHANGE OF PLASMID DNA.
CC -!- SUBCELLULAR LOCATION: Cytoplasmic.
CC -!- SIMILARITY: Belongs to the traM family.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by not-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; M20941; -; NOT_ANNOTATED_CDS.
DR PIR; A32014; A32014.
KW Conjugation; Plasmid; DNA-binding.
FT NON_TER 1
SQ SEQUENCE 5 AA; 634 MW; 6B1B1AA443500000 CRC64;

Query Match          0.4%; Score 3; DB 1; Length 5;
Best Local Similarity 100.0%; Pred. No. 1.4e+05;
Matches 3; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 333 NDE 335
DB 2 NDE 4

Search completed: August 17, 2004, 10:15:44
Job time : 14 secs

```



GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: August 17, 2004, 10:12:45 ; Search time 42 Seconds  
(without alignments)  
5529,081 Million cell updates/sec

Title: US-10-045-815-4  
Perfect score: 736  
Sequence: 1 MAVRALKLTLLAVVAAS.....PPEPAQESLSGSEKPKDEL 736

Scoring table: OLIGO  
Gapop 60.0 , Gapext 60.0

Searched: 1017041 seqs, 315518202 residues

Word size : 0

Total number of hits satisfying chosen parameters: 488

Minimum DB seq length: 0

Maximum DB seq length: 8

Post-processing: Listing first 45 summaries

Database : SPTREMBL 25:\*

- 1: sp\_archaea:\*
- 2: sp\_bacteria:\*
- 3: sp\_fungi:\*
- 4: sp\_human:\*
- 5: sp\_invertebrate:\*
- 6: sp\_mammal:\*
- 7: sp\_mbc:\*
- 8: sp\_organelle:\*
- 9: sp\_phage:\*
- 10: sp\_plant:\*
- 11: sp\_rodent:\*
- 12: sp\_virus:\*
- 13: sp\_vertebrate:\*
- 14: sp\_unclassified:\*
- 15: sp\_virus:\*
- 16: sp\_bacteriap:\*
- 17: sp\_archheap:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	4	0.5	7	Q54248	Q54248 streptomyce
2	4	0.5	8	O87471	O87471 haemophilus
3	4	0.5	8	Q51594	Q51594 escherichia
4	4	0.5	8	Q72719	Q72719 homo sapien
5	4	0.5	8	Q94623	Q94623 manduca sex
6	4	0.5	8	Q40659	Q40659 oryza sativ
7	4	0.5	8	Q42507	Q42507 triticum ae
8	4	0.5	8	P82324	P82324 pisum sativ
9	3	0.4	6	P83569	P83569 sepia offic
10	3	0.4	6	P82541	P82541 spinacia ol
11	3	0.4	7	Q8XMS3	Q8XMS3 klebsiella
12	3	0.4	7	Q47477	Q47477 escherichia
13	3	0.4	7	P72081	P72081 nocardia la
14	3	0.4	7	Q8XMS9	Q8XMS9 enterobacte
15	3	0.4	7	Q07354	Q07354 synechococc
16	3	0.4	7	P83530	P83530 lactobacill

17	3	0.4	7	4	Q8NH7	Q8nh7 homo sapien
18	3	0.4	7	5	Q7Z1C0	Q7z1c0 caenorhabdi
19	3	0.4	7	8	O9182	O9182 gnatholebia
20	3	0.4	7	8	Q95945	Q95945 saccharomyc
21	3	0.4	7	10	Q49223	Q49223 glycine max
22	3	0.4	7	10	P82445	P82445 nicotiana t
23	3	0.4	7	11	Q8K3H6	Q8k3h6 rattus norv
24	3	0.4	7	11	Q63668	Q63668 rattus norv
25	3	0.4	7	12	Q9YVE3	Q9yve3 human adeno
26	3	0.4	7	12	Q66205	Q66205 transmissib
27	3	0.4	7	12	Q66113	Q66113 cherry leaf
28	3	0.4	7	12	Q9YIQ9	Q9yiq9 human adeno
29	3	0.4	7	12	Q9YQ10	Q9yq10 transmissib
30	3	0.4	7	12	Q9YIR0	Q9yir0 human adeno
31	3	0.4	7	13	Q8JJ20	Q8jj20 gallus gall
32	3	0.4	7	13	Q42564	Q42564 fugu rubrip
33	3	0.4	8	2	Q9S6D5	Q9s6d5 escherichia
34	3	0.4	8	2	P72279	P72279 rhodococcus
35	3	0.4	8	2	Q8KPX4	Q8kpx4 microcystis
36	3	0.4	8	2	Q68485	Q68485 klebsiella
37	3	0.4	8	2	Q93SR0	Q93sr0 staphylococ
38	3	0.4	8	2	Q8RJI0	Q8rji0 streptomyce
39	3	0.4	8	2	Q3ZE29	Q3ze29 buchnera ap
40	3	0.4	8	2	Q56759	Q56759 xanthobacte
41	3	0.4	8	2	O32560	O32560 escherichia
42	3	0.4	8	2	O09258	O09258 synechococc
43	3	0.4	8	2	Q9X3K1	Q9x3k1 prochloroco
44	3	0.4	8	2	Q9ZIE9	Q9zie9 neisseria m
45	3	0.4	8	2	Q9R9E0	Q9r9e0 bacillus su

ALIGNMENTS

RESULT 1

Q54248 ID Q54248 PRELIMINARY; PRT; 7 AA.  
AC Q54248;  
DT 01-NOV-1996 (TRENBLrel. 01, Created)  
DT 01-NOV-1996 (TRENBLrel. 01, Last sequence update)  
DT 01-DEC-2001 (TRENBLrel. 19, Last annotation update)  
DE RplO protein (Fragment).  
GN RPL0.  
OS Streptomyces griseus.  
OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;  
OC Streptomycineae; Streptomycetaceae; Streptomyces.  
OX NCBI\_TaxID=1911;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=N2-3-11;  
RX MEDLINE=20011291; PubMed=10542330;  
RA Pehling S., Piepersberg W., Wehmeier U.F.;  
RT "Analysis and regulation of the sec Y gene from Streptomyces griseus  
N2-3-11 and interaction of the SecY protein with the SecA protein.";  
RL Biochim. Biophys. Acta 1447:298-302(1999).  
DR EMBL; X95915; CAA65160.1; -;  
FT NON TER  
SQ SEQUENCE 7 AA; 760 MW; 72C72B01B2D1B2A0 CRC64;

Query Match 0.5%; Score 4; DB 2; Length 7;  
Best Local Similarity 100.0%; Pred. No. 1e+06;  
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 283 VTEL 286  
Db 3 VTEL 6

RESULT 2

O87471 ID O87471 PRELIMINARY; PRT; 8 AA.  
AC O87471;  
DT 01-NOV-1998 (TRENBLrel. 08, Created)

DT 01-NOV-1998 (TReMBLrel. 08, Last sequence update)  
 DT 01-NOV-1998 (TReMBLrel. 08, Last annotation update).  
 DE HIFA (Fragment).  
 GN HIFA.  
 OS Haemophilus influenzae.  
 OC Bacteria; Proteobacteria; Gammaproteobacteria; Pasteurellales;  
 OC Pasteurellaceae; Haemophilus.  
 OX NCBI\_TaxID=727;  
 RN [1]\_TaxID=727;  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=Eagan;  
 RX MEDLINE=98389689; PubMed=97211313;  
 RA Mhlanga-Mutagadara T., Morlin G., Smith A.L., Eisenstark A.,  
 RA Golomb M.;  
 RA "Evolution of the major pilus gene cluster of haemophilus  
 RT influenzae.";  
 RL J. Bacteriol. 180:4693-4703(1998).  
 DR EMBL; AF071762; AAC35830.1; -.  
 FT NON TER 1  
 SQ SEQUENCE 8 AA; 876 MW; DAB44451A7272325 CRC64;

Query Match 0.5%; Score 4; DB 2; Length 8;  
 Best Local Similarity 100.0%; Pred. No. 1e+06;  
 Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 7 KLLT 10  
 ||||  
 Db 2 KLLT 5

RESULT 3  
 Q51594 PRELIMINARY; PRT; 8 AA.  
 AC Q51594  
 DT 01-NOV-1996 (TReMBLrel. 01, Created)  
 DT 01-NOV-1996 (TReMBLrel. 01, Last sequence update)  
 DT 01-JUN-2003 (TReMBLrel. 24, Last annotation update)  
 DE CopB protein (Fragment).  
 OS Escherichia coli.  
 OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;  
 OC Enterobacteriaceae; Escherichia.  
 OX NCBI\_TaxID=562;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=86223772; PubMed=2423502;  
 RA Weber P.C., Palchaudhuri S.;  
 RT "Incompatibility repressor in a repA-like replicon of the IncFI  
 RL plasmid ColV2-K94.";  
 RL J. Bacteriol. 166:1106-1112(1986).  
 DR EMBL; M13472; AAA23194.1; -.  
 DR GO; GO:0046821; C:extrachromosomal DNA; IEA.  
 KW Plasmid.  
 FT NON TER 1  
 SQ SEQUENCE 8 AA; 998 MW; 0F37205AA73416D7 CRC64;

Query Match 0.5%; Score 4; DB 2; Length 8;  
 Best Local Similarity 100.0%; Pred. No. 1e+06;  
 Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 486 LQRL 489  
 ||||  
 Db 1 LQRL 4

RESULT 4  
 Q72719 PRELIMINARY; PRT; 8 AA.  
 AC Q72719  
 DT 01-OCT-2003 (TReMBLrel. 25, Created)  
 DT 01-OCT-2003 (TReMBLrel. 25, Last sequence update)  
 DT 01-OCT-2003 (TReMBLrel. 25, Last annotation update)  
 DE Retinoblastoma 1 (Fragment).

OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 OX NCBI\_TaxID=9606;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RA Leone P.E., Paz-y-Mino C.S., Jervis P., Vega M.A.E., Pestana A.,  
 RA Alonso J.;  
 RA "Polymorphism of the RB1 gene in intron 16";  
 RL Submitted (MAR-2003) to the EMBL/GenBank/DBJ databases.  
 DR EMBL; AY260473; AAP39868.1; -.  
 FT NON TER 1  
 FT NON TER 8  
 SQ SEQUENCE 8 AA; 821 MW; DC65BAA72456C5A1 CRC64;

Query Match 0.5%; Score 4; DB 4; Length 8;  
 Best Local Similarity 100.0%; Pred. No. 1e+06;  
 Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 181 QNLD 184  
 ||||  
 Db 3 QNLD 6

RESULT 5  
 Q94623 PRELIMINARY; PRT; 8 AA.  
 AC Q94623  
 DT 01-FEB-1997 (TReMBLrel. 02, Created)  
 DT 01-FEB-1997 (TReMBLrel. 02, Last sequence update)  
 DT 01-NOV-1998 (TReMBLrel. 08, Last annotation update)  
 DE MsUSP-2 protein (Fragment).  
 OS USP.  
 GN Manduca sexta (Tobacco hawkmoth) (Tobacco hornworm).  
 OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;  
 OC Neoptera; Endopterygota; Lepidoptera; Glossata; Ditrysia; Spingioidea;  
 OC Spingidae; Spinginae; Manduca.  
 OX NCBI\_TaxID=71130;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=Epidermis;  
 RX MEDLINE=97165493; PubMed=9013254;  
 RA Jindra M., Huang J.Y., Malone F., Asahina M., Riddiford L.M.;  
 RT "Identification and mRNA developmental profiles of two ultraspiracle  
 RL isoforms in the epidermis and wings of Manduca sexta.";  
 RL Insect Mol. Biol. 6:41-53(1997).  
 DR EMBL; U57921; AAB64235.1; -.  
 FT NON TER 8  
 SQ SEQUENCE 8 AA; 892 MW; F165BB0415A76B16 CRC64;

Query Match 0.5%; Score 4; DB 5; Length 8;  
 Best Local Similarity 100.0%; Pred. No. 1e+06;  
 Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 290 PSRE 293  
 ||||  
 Db 3 PSRE 6

RESULT 6  
 Q40659 PRELIMINARY; PRT; 8 AA.  
 AC Q40659  
 DT 01-NOV-1996 (TReMBLrel. 01, Created)  
 DT 01-NOV-1996 (TReMBLrel. 01, Last sequence update)  
 DT 01-OCT-2002 (TReMBLrel. 22, Last annotation update)  
 DE Alpha-amylase (Fragment).  
 OS Oryza sativa (Rice).  
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;  
 OC Ehrhartoideae; Oryzae; Oryza.  
 OX NCBI\_TaxID=4530;  
 RN [1]



```

RP SEQUENCE FROM N.A.
RX MEDLINE=91078641; PubMed=2258052;
RA Kumagai M.H.; Shah M.; Terashima M.; Vrkljan Z.; Whitaker J.R.;
RA Rodriguez R.L.;
RT "Expression and secretion of rice alpha-amylase by saccharomycers
RT cerevisiae.";
RL Gene 94:209-215(1990).
DR EMBL; M62916; AAA33892.1; -.
DR Gramene; Q40659; -.
FT NON_TER 8
SQ SEQUENCE 8 AA; 948 MW; EBC694444732D6D6 CRC64;

Query Match 0.5%; Score 4; DB 10; Length 8;
Best Local Similarity 100.0%; Pred. No. 1e+06;
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 274 QVLN 277
Db ||||
2 QVLN 5

RESULT 7
Q42507 PRELIMINARY; PRT; 8 AA.
ID Q42507;
AC Q42507;
DT 01-NOV-1996 (TrEMBLrel. 01, Created)
DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
DE Heat shock protein (Fragment).
GN HSP70C OR HSP70A OR HSP70B.
OS Triticum aestivum (Wheat).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Pooideae;
OC Triticeae; Triticum.
OC NCBI_TaxID=4565;

[1]
RN SEQUENCE FROM N.A.
RP STRAIN=CV MUSTANG;
RX MEDLINE=96189275; PubMed=8605312;
RA Joshi C.P.; Kumar S.; Nguyen H.T.;
RT "Application of modified differential display technique for cloning
RT and sequencing of the 3' region from three putative members of wheat
RT HSP70 gene family.";
RL Plant Mol. Biol. 30:641-646(1996).
DR EMBL; L41507; AAB02333.1; -.
DR EMBL; L41505; AAB02331.1; -.
DR EMBL; L41506; AAB02332.1; -.
DR CO; GO:0003773; F:heat shock protein activity; IEA.
KW Heat shock.
FT NON_TER 1
SQ SEQUENCE 8 AA; 886 MW; 71B2CB1B10532768 CRC64;

Query Match 0.5%; Score 4; DB 10; Length 8;
Best Local Similarity 100.0%; Pred. No. 1e+06;
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 572 IEVV 575
Db ||||
4 IEVV 7

RESULT 8
P82324 PRELIMINARY; PRT; 8 AA.
AC P82324;
DT 01-JUN-2000 (TrEMBLrel. 14, Created)
DT 01-JUN-2000 (TrEMBLrel. 14, Last sequence update)
DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
DE Unknown protein from 2D-page of thylakoid (SPOT105) (Fragment).
OS Pisum sativum (Garden pea).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids;
OC eurosids I; Fabales; Fabaceae; Papilionoideae; Viciae; Pisum.

```

```

OX NCBI_TaxID=3888;
RN [1]
RP SEQUENCE, SUBCELLULAR LOCATION, AND DEVELOPMENTAL STAGE.
RC STRAIN=CV. DE GRACE; TISSUE=LEAF;
RX MEDLINE=20181728; PubMed=10715320;
RA Peltier J.-B.; Friso G.; Kalume D.E.; Roepstorff P.; Nilsson F.;
RA Adamska I.; van Wijk K.J.;
RT "Proteomics of the chloroplast: systematic identification and
RT targeting analysis of luminal and peripheral thylakoid proteins.";
RL Plant Cell 12:319-341(2000).
CC -!- SUBCELLULAR LOCATION: CHLOROPLAST THYLAKOID MEMBRANE LUMEN OR
CC PERIPHERY.
CC -!- DEVELOPMENTAL STAGE: UNFOLDED AND FULLY DEVELOPED LEAVES.
CC -!- MISCELLANEOUS: ON THE 2D-GEL THE DETERMINED PI OF THIS UNKNOWN
CC PROTEIN IS: 6.7, ITS MW IS: 16.8 KDA.
DR GO; GO:0009507; C:chloroplast; IEA.
DR GO; GO:0009579; C:thylakoid; IEA.
KW Chloroplast; Thylakoid.
FT NON_TER 1
FT NON_TER 8
SQ SEQUENCE 8 AA; 839 MW; DDC68B5DDDC2D2D5 CRC64;

Query Match 0.5%; Score 4; DB 10; Length 8;
Best Local Similarity 100.0%; Pred. No. 1e+06;
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 15 VVAA 18
Db ||||
2 VVAA 5

RESULT 9
P83569 PRELIMINARY; PRT; 6 AA.
AC P83569;
DT 01-JUN-2003 (TrEMBLrel. 24, Created)
DT 01-JUN-2003 (TrEMBLrel. 24, Last sequence update)
DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
DE Sperm attracting peptide SepSAF.
OS Sepia officinalis (Common cuttlefish).
OC Eukaryota; Metazoa; Mollusca; Cephalopoda; Coleoidea; Neocoleoidea;
OC Decapodiformes; Sepioidea; Sepiidae; Sepia.
OC NCBI_TaxID=6610;

[1]
RN SEQUENCE, FUNCTION, DEVELOPMENTAL STAGE, MASS SPECTROMETRY, AND
RP AMIDATION.
RX TISSUE=Egg;
RX PubMed=12207899;
RA Zatylny C.; Marvin L.; Gagnon J.; Henry J.;
RT "Fertilization in Sepia officinalis: the first mollusk sperm-
RT attracting peptide.";
RL Biochem. Biophys. Res. Commun. 296:1186-1193(2002).
CC -!- FUNCTION: ATTRACTS SPERM INCREASING THE CHANCES OF GAMETE
CC COLLISION.
CC -!- DEVELOPMENTAL STAGE: FIRST APPEARS IN THE OVARIAN FOLLICLES FURING
CC VITELLOGENESIS. ACCUMULATES IN THE OOCYTES BEFORE BEING SECRETED
CC DURING FERTILIZATION. EXPRESSION CONTINUES IN THE EMBEDDED OOCYTE.
CC ACCUMULATES IN THE EGG CAPSULE AFTER FERTILIZATION.
CC -!- MASS SPECTROMETRY: MW=556.6; METHOD=MALDI.
KW Amidation.
FT MOD_RES 6
FT MOD_RES 6
SQ SEQUENCE 6 AA; 597 MW; 72C8676AA0470000 CRC64;

Query Match 0.4%; Score 3; DB 5; Length 6;
Best Local Similarity 100.0%; Pred. No. 1e+06;
Matches 3; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 52 PGV 54
Db ||||
4 PGV 6

RESULT 10

```

P82541  
ID P82541 PRELIMINARY; PRT; 6 AA.  
AC STRAIN=CV. ALVARO; TISSUE=Leaf;  
DT 01-OCT-2000 (TRENBLrel. 15, Created)  
DT 01-OCT-2000 (TRENBLrel. 15, Last sequence update)  
DT 01-OCT-2003 (TRENBLrel. 25, Last annotation update)  
DE Chloroplast 30S ribosomal protein S19 beta (Fragment).  
OS Spinacia oleracea (Spinach).  
OG Chloroplast.  
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;  
OC Caryophyllales; Amaranthaceae; Spinacia.  
OX NCBI\_TaxID=3362;  
[1]  
RN SEQUENCE, FUNCTION, AND MASS SPECTROMETRY.  
RP STRAIN=CV. ALVARO; TISSUE=Leaf;  
RX MEDLINE=20435797; PubMed=10874039;  
RA Yamaguchi K., von Knoblauch K., Subramanian A.R.;  
RT "The plastid ribosomal proteins. Identification of all the proteins in  
the small subunit of an organelle ribosome (Chloroplast).";  
RL J. Biol. Chem. 37:28455-28465(2000).  
CC -!- FUNCTION: THIS PROTEIN BINDS DIRECTLY TO 16S RIBOSOMAL RNA.  
CC -!- SUBCELLULAR LOCATION: CHLOROPLAST.  
CC -!- TISSUE SPECIFICITY: EXPRESSED IN ALL PLANT TISSUES.  
CC -!- MASS SPECTROMETRY: MW=10477.0; METHOD=ELECTROSPRAY.  
CC -!- MASS SPECTROMETRY: MW=10495; METHOD=MALDI  
CC -!- MISCELLANEOUS: S19 ALPHA AND BETA FORMS DIFFER IN PI. S19 BETA  
FORM IS THE MINOR BASIC FORM.  
CC -!- MISCELLANEOUS: ON THE 2D-GEL ITS MW IS: 12 KDA.  
CC -!- SIMILARITY: BELONGS TO THE S19P FAMILY OF RIBOSOMAL PROTEINS.  
DR GO: GO:0009507; C:chloroplast; IEA.  
DR GO: GO:0019843; F:rRNA binding; IEA.  
DR GO: GO:0003735; P:structural constituent of ribosome; IEA.  
DR InterPro: IPR002222; Ribosomal\_S19.  
DR PROSITE: PS00323; RIBOSOMAL\_S19; PARTIAL.  
KW Ribosomal protein; Chloroplast; rRNA-binding.  
FT NON TER 6  
SQ SEQUENCE 6 AA; 732 MW; 63333735A411C000 CRC64;

Query Match 0.4%; Score 3; DB 10; Length 6;  
Best Local Similarity 100.0%; Pred. No. 1e+06;  
Matches 3; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 353 TRS 355  
DB 1 TRS 3  
[1]  
RESULT 11  
ID Q8KMS3 PRELIMINARY; PRT; 7 AA.  
AC Q8KMS3;  
DT 01-OCT-2002 (TRENBLrel. 22, Created)  
DT 01-OCT-2002 (TRENBLrel. 22, Last sequence update)  
DT 01-OCT-2002 (TRENBLrel. 22, Last annotation update)  
DE Putative MerR2 protein.  
GN MERR2.  
OS Klebsiella sp. LS13-39.  
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;  
OC Enterobacteriaceae; Klebsiella.  
OX NCBI\_TaxID=143776;  
[1]  
RN SEQUENCE FROM N.A.  
RP STRAIN=LS13-39;  
RX MEDLINE=21604134; PubMed=11763242;  
RA Mindlin S.Z., Khlodii G.Y., Gorlenko Z.M., Minakhina S.V.,  
RA Minakhin L.S., Kalyaeva E.S., Kopteva A.V., Petrova M.A.,  
RA Yurieva O.V., Nikiforov V.G.;  
RT "Mercury resistance transposons of Gram-negative environmental  
bacteria and their classification."  
RL Res. Microbiol. 152:811-822(2001).  
DR EMBL: AJ302776; CAC82975.1; -  
SQ SEQUENCE 7 AA; 608 MW; 6DC1B5BDD87DD6F0 CRC64;

Query Match 0.4%; Score 3; DB 2; Length 7;  
Best Local Similarity 100.0%; Pred. No. 1e+06;  
Matches 3; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 101 AGA 103  
DB 2 AGA 4  
[1]  
RESULT 12  
ID Q47477 PRELIMINARY; PRT; 7 AA.  
AC Q47477;  
DT 01-NOV-1996 (TRENBLrel. 01, Created)  
DT 01-MAY-1999 (TRENBLrel. 10, Last sequence update)  
DT 01-MAY-1999 (TRENBLrel. 10, Last annotation update)  
DE Tpi protein (Fragment).  
GN TPI  
OS Escherichia coli.  
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;  
OC Enterobacteriaceae; Escherichia.  
OX NCBI\_TaxID=562;  
[1]  
RN SEQUENCE OF 7-7 FROM N.A.  
RP MEDLINE=85203917; PubMed=3158524;  
RA Hellings H.W., Evans P.R.;  
RT "Nucleotide sequence and high-level expression of the major  
Escherichia coli phosphofructokinase.";  
RL Eur. J. Biochem. 149:363-373(1985).  
[2]  
RN SEQUENCE FROM N.A.  
RP Evans P.;  
RL Submitted (OCT-1986) to the EMBL/GenBank/DBJ databases.  
DR EMBL: X02519; CAA26359.1; -  
FT NON TER 1  
SQ SEQUENCE 7 AA; 773 MW; 7416D33DDDB1DB0 CRC64;

Query Match 0.4%; Score 3; DB 2; Length 7;  
Best Local Similarity 100.0%; Pred. No. 1e+06;  
Matches 3; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 229 EAA 231  
DB 2 EAA 4  
[1]  
RESULT 13  
ID P72081 PRELIMINARY; PRT; 7 AA.  
AC P72081;  
DT 01-FEB-1997 (TRENBLrel. 02, Created)  
DT 01-FEB-1997 (TRENBLrel. 02, Last sequence update)  
DT 01-DEC-2001 (TRENBLrel. 19, Last annotation update)  
DE 3'-methylcephem hydroxylase (Fragment).  
GN CEFF.  
OS Nocardia lactamdurans.  
OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;  
OC Pseudonocardiae; Pseudonocardaceae; Amycolatopsis.  
OX NCBI\_TaxID=1913;  
[1]  
RN SEQUENCE FROM N.A.  
RP MEDLINE=96009872; PubMed=7557411;  
RA Coque J., Perez-Llarena F.J., Enguita F.J., Fuente J.L., Martin J.F.,  
RA Liras P.;  
RT "Characterization of the cmcH genes of Nocardia lactamdurans and  
Streptomyces clavuligerus encoding a functional 3'-hydroxymethylcephem  
O-carbamoyltransferase for cephamycin biosynthesis.";  
RL Gene 162:21-27(1995).  
DR EMBL: Z21682; CAA79797.1; -  
FT NON TER 1  
SQ SEQUENCE 7 AA; 746 MW; 75A1B2CDD1E681C0 CRC64;

Query Match 0.4%; Score 3; DB 2; Length 7;  
Best Local Similarity 100.0%; Pred.No.1e+06; 0; Indels  
Matches 3; Conservative 0; Mismatches 0; Gaps 0;

QY 663 AVT 665  
DB 4 AVT 6

## RESULT 14

Q8KMS9 PRELIMINARY; PRT; 7 AA.  
AC Q8KMS9;  
DT 01-OCT-2002 (T-EMBLrel. 22, Created)  
DT 01-OCT-2002 (T-EMBLrel. 22, Last sequence update)  
DT 01-OCT-2002 (T-EMBLrel. 22, Last annotation update)  
DE Putative transposase (Fragment).  
GN TNIA.  
OS Enterobacter sp. CH2-4.  
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;  
OC Enterobacteriaceae; Enterobacter.  
OX NCBI\_TaxID=143777;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=CH2-4;  
RX MEDLINE=21604134; PubMed=11763242;  
RA Mindlin S.Z., Kholodii G.Y., Gorlenko Z.M., Minakhina S.V.,  
RA Minakhin L.S., Kalyaeva E.S., Kopteva A.V., Petrova M.A.,  
RA Yurieva O.V., Nikiforov V.G.;  
RT "Mercury resistance transposons of Gram-negative environmental  
bacteria and their classification."  
RL Res. Microbiol. 152:811-822(2001).  
DR EMBL; AJ302778; CAC83058.1; -.  
FT NON TER 1  
SQ SEQUENCE 7 AA; 966 MW; 737B1B1046DAA9A0 CRC64;

Query Match 0.4%; Score 3; DB 2; Length 7;  
Best Local Similarity 100.0%; Pred.No.1e+06; 0; Indels  
Matches 3; Conservative 0; Mismatches 0; Gaps 0;

QY 572 IEE 574  
DB 4 IEE 6

## RESULT 15

OC07354 PRELIMINARY; PRT; 7 AA.  
AC OC07354;  
DT 01-JUL-1997 (T-EMBLrel. 04, Created)  
DT 01-JUL-1997 (T-EMBLrel. 04, Last sequence update)  
DT 01-DEC-2001 (T-EMBLrel. 19, Last annotation update)  
DE Nifk (Fragment).  
GN NIFK.  
OS Synechococcus sp. (strain PCC 8801 / RF-1) (Cyanothecae PCC 8801).  
OC Bacteria; Cyanobacteria; Chroococcales; Cyanothecae.  
OX NCBI\_TaxID=41431;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=RF-1;  
RX MEDLINE=99231861; PubMed=10217509;  
RA Huang T.C., Lin R.F., Chu M.K., Chen H.M.;  
RT "Organization and expression of nitrogen-fixation genes in the aerobic  
nitrogen-fixing unicellular cyanobacterium Synechococcus sp. strain  
RF-1."  
RT Microbiology 145:743-753(1999).  
RL EMBL; AF003700; AAC35193.1; -.  
FT NON TER 1  
SQ SEQUENCE 7 AA; 849 MW; 7412C72AA9D5B030 CRC64;

Query Match 0.4%; Score 3; DB 2; Length 7;  
Best Local Similarity 100.0%; Pred.No.1e+06;  
Matches 3; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 692 DLV 694  
DB 4 DLV 6

Search completed: August 17, 2004, 10:16:38  
Job time : 45 secs

This Page Blank (uspto)